AD_____

GRANT NUMBER DAMD17-97-1-7118

TITLE: Few-View Tomographic Reconstruction of Technetium-99m-
sestamibi Distribution for the Detection and Differentation of
Breast Lesions

PRINCIPAL INVESTIGATOR: Patrick J. La Riviere
                          Chin-Tu Chen

CONTRACTING ORGANIZATION: The University of Chicago
                            Chicago, Illinois 60637

REPORT DATE: August 1999

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
               Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for public release;
                          distribution unlimited

The views, opinions and/or findings contained in this report are
those of the author(s) and should not be construed as an official
Department of the Army position, policy or decision unless so
designated by other documentation.

20000828 117

REPORT DOCUMENTATION PAGE

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE August 1999	3. REPORT TYPE AND DATES COVERED Annual (15 Jul 98 –14 Jul 99)

4. TITLE AND SUBTITLE
Few-View Tomographic Reconstruction of Technetium-99m-sestamibi Distribution for the Detection and Differentation of Breast Lesions

5. FUNDING NUMBERS
DAMD17-97-1-7118

6. AUTHOR(S)

Patrick J. La Riviere
Chin-Tu Chen

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

The University of Chicago
Chicago, IL 60637

E*Mail: pjlarivi@midway.uchicago.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The broad objective of the research is to develop and evaluate methods for the reconstruction of dedicated breast single-photon emission computed tomography (SPECT) scintimammography (SMM) images from a relatively small number of projection views. We have pursued a strategy involving sinogram preprocessing, in which each projection view is first smoothed using Fourier or spline-based techniques and then additional projection views are interpolated, again using Fourier or spline-based techniques, prior to reconstruction by filtered backprojection (FBP). The spline-based projection smoothing technique is a novel application of roughness-penalized nonparametric regression using an explicit Poisson model and is found to be superior to traditional Fourier smoothing techniques. Likewise, the spline-based projection interpolation approach is found to be more accurate in practical situations than the Fourier-based approach. When applied in concert, the two spline-based techniques yield high-quality images from as few as 15 projection views. We have also developed two alternatives to reconstruction by FBP, one based on the direct Fourier method, the other on direct reconstruction from the coefficients of splines fit to the projection views.

14. SUBJECT TERMS
Breast Cancer, scintimammograph, tomography, SPECT, Tc-99m-sestamibi, interpolation, smoothing

15. NUMBER OF PAGES
96

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

_____ Where copyrighted material is quoted, permission has been obtained to use such material.

_____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

_____ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

_____ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

__/__ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

_____ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

_____ In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

_____ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.
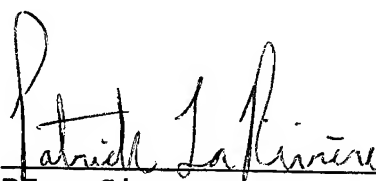
Patrick LaRiviere   8 August, 1999
PI - Signature   Date

Table of Contents

1 Introduction

The broad objective of the proposed research is to develop, implement, and evaluate methods for the reconstruction of dedicated single-photon emission computed tomography (SPECT) scintimammography (SMM) images from a relatively small number of projection views. SMM is a nuclear-medicine test with the potential to provide relatively low-cost, minimally invasive differentiation of breast abnormalities identified by physical examination or mammography. It relies on the preferential uptake of Tc-99m-sestamibi or other radionuclides in breast malignancies as compared to normal breast tissue or benign abnormalities. This focal uptake can be imaged in a number of ways, though the most widely used clinical protocol involves acquiring one or two planar views while the patient lies prone on a specially designed table. However, preliminary experimental work, verified theoretically in our own work, has suggested that a dedicated breast SPECT geometry, in which a small camera revolves around a dependent breast, would provide better lesion detectability than do the planar or conventional SPECT geometries. The drawback of this approach would be the relatively long time needed to acquire the number of projection views traditionally used for SPECT image reconstruction. The aim of this research, then, is to reduce imaging time in dedicated breast SPECT by developing algorithms that allow images to be reconstructed from a smaller number of projection views than is conventionally used while maintaining diagnostically useful image quality. Whereas the original strategy was to develop iterative reconstruction techniques incorporating prior assumptions about the structures being imaged, these assumptions proved difficult to quantify, the resulting algorithms extremely computationally intensive, and the reconstructed images relatively poor. We have thus pursued an alternative strategy involving sinogram preprocessing, in which each projection view is first smoothed using Fourier or spline-based techniques and then additional projection views are interpolated, again using Fourier or spline-based techniques, prior to reconstruction by filtered backprojection (FBP).

2 Body[1]

The research accomplishments to date can be grouped naturally into 5 categories. The first was the comparison of SMM imaging geometries—planar, conventional SPECT, and dedicated breast SPECT—to determine which offered the best lesion detectability. The second was the investigation of iterative reconstruction techniques for few-view tomography. Given the shortcomings that became evident in examining these approaches, we then turned to a strategy of sinogram preprocessing prior to reconstruction by FBP, which comprised the next two accomplishments: the evaluation of methods for the interpolation of additional projection views and the development of novel projection-smoothing techniques. The final accomplishment involved the development of two alternatives to FBP reconstruction that emerged naturally from the sinogram preprocessing techniques.

2.1 Comparison of imaging geometries

The research began with verification of the hypothesis that a dedicated breast SPECT geometry would indeed be better for SMM lesion detection than currently existing planar or conventional SPECT geometries. The experimental results of Wang *et al.* [1] suggested that it would be, but we undertook to answer the question more definitively and quantitatively by using the so-called ideal-observer framework to quantify the amount of information contained in the projections of a breast phantom using the three different geometries. The results indicated conclusively that the dedicated breast SPECT geometry provides better lesion detectability than the other two [2].

2.2 Investigation of iterative reconstruction techniques

After establishing the superiority of the dedicated breast SPECT geometry, we turned to the question of developing iterative reconstruction algorithms tailored to few-view SMM data. We implemented the standard algebraic recon-

---

[1] In accordance with the document *(Annual Summary) Training Reporting Requirements,* this report summarizes research performed "to date." Sections of this report detailing work performed primarily in the second year of the grant (July 15, 1998—July 14, 1999) are designated by a *.

struction technique (ART) algorithm, with disappointing results. Even in the absence of noise, the quality of the reconstructed images was not sufficiently high, as compared to FBP, to justify the large computational burden of the approach. In the presence of noise, images reconstructed by the ART approach were quite poor, even compared to those reconstructed by ramp-filtered FBP.

Attempts to improve the quality of the reconstructions by incorporating explicit statistical information as well as prior assumptions about the simplicity, symmetry, and near-binary nature of the expected reconstructions met with numerous difficulties. The most natural constraints did not lend themselves to efficiently implementable mathematical formulations and approximations of the constraints led to computationally intensive algorithms with little apparent advantage over FBP. These shortcomings of the iterative approaches led us to focus on more computationally efficient sinogram preprocessing strategies, in which each projection view is first smoothed using Fourier or spline-based techniques and then additional projection views are interpolated, again using Fourier or spline-based techniques, prior to reconstruction by filtered backprojection (FBP).

2.3   Interpolation of additional projections prior to reconstruction by FBP

The minimum number of angular views required to produce an accurate tomographic reconstruction of a given object using FBP is dictated by two factors. First, the angular sampling of the object's sinogram must satisfy, at least approximately, the Nyquist sampling condition. Absent this, any reconstruction is doomed to suffer from angular aliasing artifacts. Second, the number of angular samples must satisfy FBP's implicit assumptions about the density of angular sampling. It is well known that FBP reconstructions from a small number of angular views are degraded by prominent star-shaped artifacts.

The Nyquist condition is the more fundamental of these two sampling requirements, because when it is satisfied by a number of samples less than the number required by the reconstruction algorithm, it is in principle possible to interpolate exactly the additional views needed. This will often be the case in SPECT SMM, owing to the relative simplicity and symmetry of the object being imaged. Ideally, a periodic interpolation method should be chosen in order to make use of the inherent periodicity of the angular samples.

We have examined the use of two such methods: Fourier-based interpolation by means of fast Fourier transform (FFT) zero-padding and periodic spline interpolation.

2.3.1   Fourier-based interpolation

Fourier-based interpolation by means of FFT zero-padding entails taking the FFT of a sequence of samples and extending the result with zeros. Taking an inverse FFT of the padded sequence yields a more densely sampled version of the original sequence. In the case of few-view tomography, we apply this technique to the sinogram angular samples at each projection bin independently, thereby increasing the number of projection views. While this technique may seem somewhat crude, it can, in fact, be shown that is exact for periodic functions that are sampled in accordance with the Nyquist condition [3]. This sampling condition is often met to good approximation in SPECT SMM for relatively small ($\sim$30) numbers of projection views.

We have also shown that Fourier-based interpolation has favorable noise properties [4]. Specifically, when interpolating among samples contaminated with zero-mean white noise of a given variance, the resulting interpolated curve has stationary noise with the same variance. When interpolating among samples contaminated with Poisson noise, the resulting interpolated curve has variance that is flat locally and that tracks the noise in the measured samples over longer distances. These properties are desirable because wide fluctuations in noise levels in the interpolated curve can lead to artifacts in reconstructed images.

Despite these favorable theoretical properties, Fourier-based interpolation can be sensitive to violations of the Nyquist sampling condition. For this reason, we have explored a second, spline-based approach that may prove more robust in practical situations.

6

2.3.2 Spline-based interpolation

A spline is a piecewise polynomial curve that has many desirable numerical and statistical properties. In few-view tomography, we fit a spline satisfying periodic boundary conditions to the angular samples at each projection bin using computationally efficient algorithms and then resample the splines to obtain additional projection views. We have derived the noise properties for periodic spline interpolation and found them to be favorable [4]. Specifically, when interpolating among samples contaminated with zero-mean white noise of a given variance, the resulting interpolated curve displays only a slight dip in noise level between the positions of the measured samples. Similar local behavior is observed in the case when the samples are contaminated by Poisson noise.

2.3.3 Comparison of accuracy of spline, Fourier, and linear interpolation of additional projections

We have compared the accuracy of the zero-padding and periodic spline interpolation approaches, as well as linear interpolation with periodic boundary conditions, for the task of interest: interpolating additional projections in a few-view sinogram [5]. Simply comparing the success of the approaches in interpolating a single sinogram each for one or two canonical phantoms would have provided more anecdotal than genuinely rigorous evidence on which to base the choice of interpolation method for few-view tomography. The outcome could have depended as much on numerical happenstance as on the genuine strengths of the approaches. Instead, we generated 100 different "realizations" of each of the two types of numerical phantom—Shepp-Logan and breast—by choosing the parameters specifying the constituent ellipses of each type to vary according to predetermined probability laws. Corresponding sinograms of 128 bins x 1024 projection views were computed analytically and subsampled to 16, 32, 64, 128, 256, and 512 projection views. Each subsampled sinogram was interpolated to 1024 projection views by each of the methods under consideration and the normalized root-mean-square-error (NRMSE) with respect to the true 1024 projection view sinogram computed. In addition, images were reconstructed from the interpolated sinograms by FBP and the NRMSE with respect to the true phantom computed. The non-parametric signed rank test was then used to assess the statistical significance of the pairwise differences in mean NRMSE among the interpolation methods for the various conditions: phantom family (Shepp-logan or breast), number of measured projection views (16, 32, 64, 128, 256, or 512), and endpoint (sinogram or image). Periodic spline interpolation was found to be superior to the others in a statistically significant way for virtually every condition.

2.4 Novel projection smoothing methods

Because few-view tomography images are reconstructed using fewer total counts than are images reconstructed from a standard number of views, careful noise control is essential. In this situation, the standard tomographic practice of mitigating noise by multiplying the Fourier transform of each projection by an apodization window may not be sufficient. This is equivalent to a shift-invariant smoothing of each projection, whereas a shift-variant smoothing that more accurately accounted for the measurement statistics might be preferable.

2.4.1 *Spline-based nonparametric regression

In this vein, we have investigated a projection smoothing technique based on roughness-penalized nonparametric regression using an explicit Poisson model [6]. This is different from, and better than, the smoothing approach described in last year's report, which entailed smoothing Fourier coefficients and did not explicitly account for the Poisson nature of the data. In the new approach, each set of projection samples $y_i = p^{(\phi_j)}(\xi_i)$, $i = 0, ..., N - 1$, where ϕ_j denotes the projection view, ξ_i the projection bin, and N the number of bins per projection, is fit with a curve $\widehat{p}^{(\phi_j)}(\xi_i)$ maximizing a roughness-penalized Poisson likelihood function

$$\Phi(p, y) = \sum_{i=0}^{N-1} [y_i \ln p(\xi_i) - p(\xi_i)] - \alpha \int [p''(\xi)]^2 \, d\xi, \tag{1}$$

where $''$ denotes the second derivative, α is the smoothing parameter, and where, for simplicity, we have dropped the dependence on ϕ_j. It can be shown that this objective function is always maximized by a natural cubic spline, which is simply a cubic spline constrained to be linear beyond the first and last measurement points. The coefficients of the spline can be found through a computationally efficient iterative procedure. This approach captures some of the statistical benefits of a fully iterative reconstruction algorithm, although because it operates on each projection independently, the computational cost is a fraction of that of a fully iterative algorithm.

2.4.2 *Comparison of spline-based projection smoothing with Fourier-based apodization window smoothing

We have compared the roughness-penalized nonparametric regression smoothing approach with standard Fourier-based apodization window smoothing by computing resolution-variance curves. Because the roughness-penalized nonparametric regression smoothing is not shift invariant, the resolution in the reconstructed images is not uniform, so resolution-variance curves were computed at a variety of points for a typical phantom. In all cases, the nonparametric regression smoothing curves were lower than those for the apodization window, indicating superior performance.

We have also investigated the nonuniform resolution induced in the images by use of the novel projection smoothing technique and found that it can be controlled by adjusting the so-called *link function* of the model so as to yield the favorable outcomes of having better resolution in higher-count areas than in lower-count areas or of having essentially uniform resolution.

2.4.3 *Evaluation of combined nonparametric regression and spline-based interpolation for few-view tomography

Because of the strong performance of the spline-based smoothing and interpolation approaches, we applied them in concert to the reconstruction of few-view emission tomography images [6]. In order to examine the response of the algorithm to increasingly difficult interpolation tasks, we imaged a compact physical phantom that could be placed at increasing radial offsets in a three-headed SPECT system. This effectively increased the bandwidth of the angular functions at each projection bin. Specifically, we acquired projections of a Data Spectrum ventricular phantom placed at five different radial offsets from the center of rotation: 0, 5, 9, 12, and 15 cm. The phantom contained a 1-cm defect insert. From this data we extracted 3D sinograms corresponding to 15, 30, 60, and 120 views, respectively. Thus we had 20 different sinograms, corresponding to the 20 possible combinations of radial offset and number of angular views. We reconstructed images from these 20 sinograms using four different processing techniques:

1. No pre-smoothing of the sinogram and reconstruction from available views by FBP using a Hanning filter.

2. No pre-smoothing of the sinogram, spline interpolation from the available views to 120 angular views, and reconstruction by FBP using a Hanning filter.

3. Roughness-penalized nonparametric regression smoothing of the sinogram and reconstruction from the available views by FBP using a ramp filter.

4. Roughness-penalized nonparametric regression smoothing of the sinogram, spline interpolation from the available views to 120 views, and reconstruction by FBP using a ramp filter.

We observed that reconstructions from available views without pre-smoothing or interpolation displayed star-shaped artifacts and a mottled appearance when the number of views was small. Interpolation alone mitigated the star-shaped artifacts but lead to severe circular artifacts, particularly in the case of a small number of views and a large radial offset. Smoothing alone reduced the noise visibility but had little effect on the star-shaped artifacts. The combination of smoothing and interpolation, however, led to visually appealing reconstructions for most combinations of radial offset and number of views, including as few as 15 angles in the 0-cm offset case. Generation of bullseye plots indicated that the defect insert remained detectable in the reconstructed images for all but the most extreme combinations of radial offset and number of views.

2.5 Novel direct Fourier and spline-based reconstruction algorithms

While the sinograms processed using the Fourier- and spline-based techniques can be reconstructed by FBP, we have also developed and investigated two additional reconstruction approaches that more naturally exploit the strengths of these processing techniques.

2.5.1 *Development of a direct Fourier reconstruction method

While the Fourier interpolation technique discussed above was used to increase the number of projection views in the sinogram prior to reconstruction by FBP, it can also be used in the implementation of a *direct Fourier* reconstruction technique. Direct Fourier techniques are based on the central slice theorem, which states that the Fourier transform of each projection view corresponds to a line through the origin in the Fourier transform space of the object being imaged. The Fourier transforms of a set of projection views then provides a set of polar samples of the object's Fourier transform. If the samples are interpolated onto a Cartesian grid, the object can then be reconstructed by use of the FFT. Our strategy is to use zero-padding interpolation to increase the density of polar samples in both the radial and azimuthal directions, after which linear interpolation is used to generate the Cartesian samples. We have found this approach to be both accurate and computationally efficient [3].

2.5.2 Development of a direct spline reconstruction method

When smoothing splines are fit to each projection in the nonparametric regression technique, they must be resampled to yield a discrete sinogram if reconstruction is to proceed by FBP. Discarding the continuous information embodied in the spline coefficients seems wasteful because after filtration of these samples the FBP algorithm interpolates among them during backprojection. We have an investigated an alternative to FBP in which the reconstructed image is expressed explicitly in terms of the coefficients of the splines fit to the projections. While evaluating this expression is computationally intensive, the approach is found to yield higher resolution images than does FBP [7].

References

[1] H. Wang, C. Scarfone, K. L. Greer, R. E. Coleman, and R. J. Jaszczak, "Prone breast tumor imaging using vertical axis-of-rotation SPECT systems: An initial study," *IEEE Trans. Nucl. Sci.*, vol. 44, pp. 1271–1276, 1997.

[2] P. J. La Rivière, X. Pan, and B. C. Penney, "Ideal-observer analysis of lesion detectability in planar, conventional SPECT, and dedicated SPECT scintimammography using effective multi-dimensional smoothing," *IEEE Trans. Nucl. Sci.*, vol. 45, pp. 1273–1279, 1998.

[3] P. J. La Rivière and X. Pan, "Mathematical equivalence of zero-padding and circular sampling thorem interpolation with implications for direct Fourier image reconstruction," in *Proc. SPIE*, vol. 3338, pp. 1117–1126, 1998.

[4] P. J. La Rivière and X. Pan, "Noise properties of periodic interpolation methods with implications for few-view tomography," *IEEE Trans. Nucl. Sci.*, vol. 46, pp. 639–645, 1999.

[5] P. J. La Rivière and X. Pan, "Comparison of angular interpolation approaches in few-view tomography using statistical hypothesis testing," in *Proc. SPIE*, vol. 3661, pp. 398–407, 1999.

[6] P. J. La Rivière and X. Pan, "Few-view tomography using roughness-penalized nonparametric regression and periodic spline interpolation," *IEEE Trans. Nucl. Sci.*, vol. 46, 1999. (in press).

[7] P. J. La Rivière and X. Pan, "Spline-based inverse Radon transform in two and three dimensions," *IEEE Trans. Nucl. Sci.*, vol. 45, pp. 2224–2231, 1998.

Appendix 1: Key research accomplishments

- We have determined, through use of the ideal-observer framework, that the dedicated breast SPECT geometry is superior to the planar or conventional SPECT geometries for scintimammography.

- We have developed a Fourier-based interpolation technique for increasing the number of projection views in few-view tomography.

- We have analyzed the basic resolution and noise properties of Fourier-based interpolation.

- We have developed a spline-based interpolation technique for increasing the number of projection views in few-view tomography.

- We have analyzed the basic resolution and noise properties of spline-based interpolation.

- We have compared the accuracy of Fourier, spline, and linear interpolation for increasing the number of projection views in few-view tomography and found spline interpolation to be the most accurate in general.

- We have developed a novel projection-smoothing technique based on roughness-penalized nonparametric regression using an explicit Poisson noise model.

- We have compared this novel projection-smoothing technique to conventional Fourier-domain apodization window techniques by computing resolution-variance curves and found the novel approach to be superior.

- We have found that the nonuniform resolution induced in the images by use of the novel projection smoothing technique can be controlled by adjusting the so-called *link function* of the model to yield the favorable outcomes of having better resolution in higher-count areas than in lower-count areas or of having essentially uniform resolution.

- As an alternative to FBP, we have investigated the use of direct Fourier reconstruction, in which zero-padding interpolation is used to increase the density of polar samples in the Fourier transform space of the object being imaged, after which linear interpolation is used to estimate Cartesian samples for reconstruction by the 2D FFT. We have found this approach to be both accurate and computationally efficient.

- As a second alternative to FBP, for use when smoothing splines are fit to the projections, we have investigated a direct-spline reconstruction technique in which the reconstructed image is expressed in terms of the coefficients of the splines. The technique is found to yield image resolution superior to that of FBP, but at considerable computational cost.

Appendix 2: Reportable outcomes

Peer-Reviewed Original Articles

1. La Rivière, P. J. and Pan, X., "Noise properties of periodic interpolation methods with implications for few-view tomography," *IEEE Trans. Nucl. Sci.,* 1999, **46**, pp. 639–645, 1999.

2. La Rivière, P. J. and Pan, X., "Few-view tomography using roughness penalized nonparametric regression and periodic spline interpolation," *IEEE Trans. Nucl. Sci.,* 1999 (in press).

3. Kao, C.-M., Pan, X., Anastasio, M., and La Rivière, P. J., "A Fourier-based optimal recovery approach for anti-aliasing interpolation," *Optical Engineering,* 1999 (in press).

4. La Rivière, P. J., Pan, X., and Penney, B. C., "Ideal-observer analysis of lesion detectability in planar, conventional SPECT, and dedicated SPECT scintimammography using effective multi-dimensional smoothing," *IEEE Trans. Nucl. Sci.,* **45**, pp. 1273-1279, 1998.

5. La Rivière, P. J. and Pan, X. "Spline-based inverse Radon transform in two and three dimensions," *IEEE Trans. Nucl. Sci.,* **45**, pp. 2224-2231, 1998.

Proceedings Articles

1. La Rivière, P. J. and Pan, X., "Few-view tomography using interpolating and smoothing splines with implications for cardiac SPECT," *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.,* **3**, pp. 1615-1619, 1999.

2. La Rivière, P. J. and Pan, X., "Noise properties of periodic interpolation methods with implications for few-view tomography," *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.,* **3**, pp. 1610-1614, 1999.

3. La Rivière, P. J. and Pan, X., "Comparison of angular interpolation approaches in few-view tomography using statistical hypothesis testing," *Proc. SPIE,* **3661,** pp. 398–407, 1999.

4. La Rivière, P. J., Pan, X., Penney, B. C., and C.-T., Chen, "Improved detectability of malignant lesions in SPECT scintimammography using effective multi-dimensional smoothing," *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.,* **2**, pp. 1581-1585, 1998.

5. La Rivière, P. J. and Pan, X., "Direct spline-based inversion of the three-dimensional Radon transform with application to cardiac phantom data," *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.,* **2**, pp. 1674-1679, 1998.

6. La Rivière, P. J. and Pan, X., "Mathematical equivalence of zero-padding interpolation and circular sampling theorem interpolation with implications for direct Fourier image reconstruction," *Proc. SPIE,* **3338,** pp. 1117-1126, 1998.

7. Pan, X., La Rivière, P. J., Ye, J., Mukherjee, J. and Chen, C.-T., "Efficient sinogram smoothing for dynamic neuroreceptor PET imaging," *Proc. SPIE,* **3033,** pp. 140-146, 1997.

Presentations

1. La Rivière, P. J. and Pan, X., "Comparison of angular interpolation approaches for few-view tomography using statistical hypothesis testing," *SPIE Medical Imaging Conference,* 1999.

2. La Rivière, P. J. and Pan, X., "Noise properties of periodic interpolation methods with implications for few-view tomography," *IEEE Medical Imaging Conference,* 1998.

3. Kao, C.-M., Pan, X., Anastasio, M., and La Rivière, P. J., "An interpolation method using signal recovery and discrete Fourier transform," *IEEE Medical Imaging Conference,* 1998.

4. La Rivière, P. J., Pan, X., and Penney B. C., "Few-view tomography using interpolating and smoothing splines with implications for cardiac SPECT," *IEEE Medical Imaging Conference,* 1998.

5. La Rivière, P. J. and Pan, X., "Mathematical equivalence of zero-padding interpolation and circular sampling theorem interpolation with implications for direct Fourier image reconstruction," *SPIE Medical Imaging Conference,* 1998.

6. La Rivière, P. J. and Pan, X., "Direct spline-based inversion of the three-dimensional Radon transform with application to cardiac phantom data," *IEEE Medical Imaging Conference,* 1997.

7. La Rivière, P. J. and Pan, X., "Improved detectability of malignant lesions in SPECT scintimammography using effective multi-dimensional smoothing, *IEEE Medical Imaging Conference,* 1997.

8. La Rivière, P. J., Pan, X., Penney, B. C., Chen, C.-T., and Reba, R., "Few-angle cardiac SPECT with spline processing: initial results and radial-extent limitations," *83rd Scientific Assembly and Annual Meeting of Radiological Society of North America*, 1997.

9. La Rivière, P. J. and Pan, X., "Direct spline-based inversion of the three-dimensional Radon transform with application to cardiac SPECT," *83rd Scientific Assembly and Annual Meeting of Radiological Society of North America*, 1997.

10. La Rivière, P. J., Pan, X., Penney, B. C., and Chen, C.-T., "Improved detectability of malignant lesions in dedicated SPECT scintimammography using effective multi-dimensional smoothing," *83rd Scientific Assembly and Annual Meeting of Radiological Society of North America,* 1997.

Accepted Presentations

1. La Rivière, P. J. and Pan, X., "Resolution properties of non-parametric regression sinogram smoothing using an explicit Poisson model," accepted as an oral presentation for the *IEEE Medical Imaging Conference,* 1999.

2. La Rivière, P. J., Pan, X., and Kao, C.-M., "Medical imaging applications of a novel multi-dimensional interpolation approach," accepted as a poster presentation for the *IEEE Medical Imaging Conference,* 1999.

Appendix 3: Attached articles

The peer-reviewed and proceedings articles listed in appendix 2 are appended following this page.

Noise Properties of Periodic Interpolation Methods with Implications for Few-View Tomography

P. J. La Rivière, *Student Member, IEEE,* and X. Pan, *Member, IEEE*

Department of Radiology, The University of Chicago, Chicago, IL, 60637

Abstract

A number of methods exist specifically for the interpolation of periodic functions from a finite number of samples. When the samples are known exactly, exact interpolation is possible under certain conditions, such as when the function is bandlimited to the Nyquist frequency of the samples. However, when the samples are corrupted by noise, it is just as important to consider the noise properties of the resulting interpolated curve as it is to consider its accuracy. In this work, we derive analytic expressions for the covariance and variance of curves interpolated by three periodic interpolation methods—circular sampling theorem, zero-padding, and periodic spline interpolation—when the samples are corrupted by noise. We perform empirical studies for the special cases of white and Poisson noise and find the results to be in agreement with the analytic derivations. The implications of these findings for few-view tomography are also discussed.

I. INTRODUCTION

The need to interpolate samples of periodic functions arises in a number of important medical imaging applications. For instance, when performing emission computed tomography imaging of a compact, reasonably symmetric object, such as the breast, one can achieve adequate angular sampling of the object's sinogram with a relatively small number of projection views. However, using filtered backprojection (FBP) to reconstruct the image may still lead to star-shaped artifacts, because FBP implicitly requires a relatively high density of angular samples [1]. In these situations, periodic interpolation may be used to interpolate additional angular views between the measured ones in order to satisfy FBP's sampling requirements. The need to perform periodic interpolation also arises in direct Fourier image reconstruction, where it is necessary to interpolate from a polar to a Cartesian grid in Fourier space [2]. The interpolation is often accomplished using separate 1D interpolations in the radial and azimuthal directions, and the azimuthal interpolation should rightly be periodic.

There exist a number of methods for interpolating periodic functions. Circular sampling theorem (CST) interpolation, for one, is a special case of Whittaker-Shannon (W-S) sinc interpolation that applies to periodic functions [3, 4]. Consider a periodic function $g(x)$ that has period X and which is bandlimited to frequency K (i.e., the coefficients of expansion a_k of the function's Fourier series satisfy $a_k = 0$ for $|k| > K$). Given $N \geq 2K + 1$ samples of $g(x)$ taken at points $x_n = nX/N$ ($n = 0, \ldots, N - 1$) evenly spaced over one period, the CST states [4] that $g(x)$ can be interpolated exactly

by use of

$$g(x) = \sum_{n=0}^{N-1} g(x_n)\sigma_N(x - x_n), \tag{1}$$

where

$$\sigma_N(x) = \sin[(2K + 1)\pi x/X]/N \sin(\pi x/X). \tag{2}$$

If the Nyquist condition is not satisfied, that is, if $g(x)$ is not truly bandlimited to frequency K or if $N < 2K + 1$, Eqs. 1 and 2 no longer represent exact interpolation, but they remain mathematically meaningful and, often, practically useful. For instance, if the spectral components beyond frequency K are negligibly small but not exactly zero, interpolating with Eqs. 1 and 2 remains very accurate.

A second periodic interpolation approach, zero-padding (ZP) interpolation, involves extending the discrete Fourier transform (DFT) of a finite sequence with zeroes and taking an inverse DFT to generate a more densely sampled version of the original sequence with values interpolated at intermediate positions between the original measured samples [5–8]. Specifically, one begins by taking the DFT of the sequence $g(x_n)$, which is given by

$$c_k = \frac{1}{N} \sum_{n=0}^{N-1} g(x_n) \exp(-j2\pi nk/N), \tag{3}$$

for $k = 0, \ldots, N - 1$, where $j = \sqrt{-1}$. Zero-padding involves the creation of a new sequence $d_{k'}$, having $L = PN$ elements (where P is an integer). If $g(x)$ is assumed to be bandlimited to frequency K and if $N \geq 2K + 1$, the sequence $d_{k'}$ is defined as follows:

$$d_{k'} = \begin{cases} c_{k'} & k' = 0, \ldots, K \\ 0 & k' = K + 1, \ldots, L - K - 1 \\ c_{k'-L+N} & k' = L - K, \ldots, L - 1. \end{cases} \tag{4}$$

A more densely sampled sequence $g(x_l)$, where $x_l = lX/L$ ($l = 0, \ldots, L - 1$), is now obtained by taking the inverse DFT of the sequence $d_{k'}$,

$$g(x_l) = \sum_{k'=0}^{L-1} d_{k'} \exp(j2\pi lk'/L), \tag{5}$$

for $l = 0, \ldots, L - 1$. ZP interpolation is generally viewed as a somewhat crude interpolation approach, but in fact nothing could be further from the truth. It can be shown that ZP interpolation is equivalent to CST interpolation, in that the

spatial-domain interpolation function corresponding to the ZP operation in frequency space is just $\sigma_N(x)$ given in Eq. 2 [9]. That is, it can be shown that

$$g(x_l) = \sum_{n=0}^{N-1} g(x_n)\sigma_N(x_l - x_n), \qquad (6)$$

where $\sigma_N(x)$ is given by Eq. 2. This equivalence holds regardless of whether the Nyquist condition is satisfied. If it is satisfied, then ZP interpolation, like CST interpolation, is exact. Obviously, the CST interpolation formula can be used to estimate $g(x)$ at any arbitrary point x whereas zero-padding interpolation is constrained to interpolate onto a fixed, equispaced grid P times denser than the original samples. However, since ZP yields a continuous interpolated curve in the limit as $P \to \infty$, we shall treat these two approaches as one in the subsequent analysis, using the continuous form of Eq. 1 to represent them both.

The final periodic interpolation method to be examined is periodic spline (PS) interpolation. Splines are piecewise cubic polynomials that are continuous up to and including the second derivative at the joints between pieces [10, 11]. Periodic splines are further constrained to satisfy periodic boundary conditions. A spline $\widehat{g}(x)$ can be represented by

$$\widehat{g}(x) = a_n + b_n(x - x_n) + c_n(x - x_n)^2 + d_n(x - x_n)^3, \qquad (7)$$

for $x \in [x_n, x_{n+1}]$, where the x_n are the abscissas at which the data is measured and $n = 0, \ldots, N - 1$. Fitting a spline is thus tantamount to finding the coefficients a_n, b_n, c_n, and d_n subject to interpolation, continuity, and boundary conditions. Because evaluating a spline at a particular point x involves evaluating Eq. 7 using the a_n, b_n, c_n, and d_n corresponding to the interval $[x_n, x_{n+1}]$ in which x falls, we can think of representing a spline as an N-component vector of functions, in which the nth component corresponds to the interval $[x_n, x_{n+1}]$ and contains a function of the form of Eq. 7 with the appropriate values substituted for a_n, b_n, c_n, and d_n. To reflect this understanding and simplify later calculations we introduce the somewhat unconventional notation

$$\widehat{\mathbf{g}}(x) = \mathbf{a} + \mathcal{D}(x)\mathbf{b} + \mathcal{D}^2(x)\mathbf{c} + \mathcal{D}^3(x)\mathbf{d}, \qquad (8)$$

where $\widehat{\mathbf{g}}(x)$ is an N-element vector of functions, $\mathbf{a}$ is the N-element vector with coefficients a_n and likewise for $\mathbf{b}$, $\mathbf{c}$, and $\mathbf{d}$, and $\mathcal{D}$ is a diagonal matrix with $\mathcal{D}_{nn} = x - x_n$.

In fitting a spline, the coefficients are obtained by linear operations on the measured data. If the measured samples $g(x_n)$ are represented as an N-element vector $\mathbf{g}$, the vectors of coefficients can be found from $\mathbf{g}$ through matrix multiplications $\mathbf{a} = A\mathbf{g}$, $\mathbf{b} = B\mathbf{g}$, $\mathbf{c} = C\mathbf{g}$, and $\mathbf{d} = D\mathbf{g}$, where the matrices A, B, C, and D can be deduced from [11]. Substituting for $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, and $\mathbf{d}$ in (8) in terms of these matrix products yields

$$\widehat{\mathbf{g}}(x) = \left[A + \mathcal{D}(x)B + \mathcal{D}^2(x)C + \mathcal{D}^3(x)D\right]\mathbf{g}. \qquad (9)$$

While not theoretically exact for the interpolation of periodic, bandlimited functions, periodic splines can nonetheless be very accurate in that situation [12]. And in practice, when interpolating functions that are not exactly bandlimited, periodic splines often outperform CST and ZP interpolation, which are quite sensitive to departures from the bandlimited assumption, giving rise to high-frequency (Gibbs) artifacts [13]. In fact, if the smoothness of a function is defined in terms of the function's integrated second derivative, it can be shown that splines are the smoothest possible interpolant to any set of samples [14].

While questions about the accuracy of various interpolation approaches are paramount when the measured samples are known exactly, other concerns arise when the samples are known to be corrupted by noise. In particular, it becomes important to analyze how the noise is propagated into the interpolated samples. Consider now that the samples of $g(x)$ are corrupted by additive, zero-mean noise. These noisy samples can be represented as

$$g(x_n) = \langle g(x_n) \rangle + n(x_n). \qquad (10)$$

where $g(x_n)$ and $n(x_n)$ are random variables, the latter representing zero-mean additive noise, and $\langle \rangle$ represents the expectation operator. The aim of this paper is to derive the covariance and variance of the curves interpolated by means of CST/ZP interpolation and PS interpolation, to compare these analytic predictions with results of Monte Carlo simulations, and to draw conclusions from these results about the suitability of the various approaches for the interpolation task encountered in few-view tomography.

II. METHODS

A. Analytic Derivations

Let $\widehat{g}(x)$ be a curve interpolated from the noisy samples of Eq. 10 by CST or ZP interpolation. The covariance between two points x and x' of this function is given by

$$\mathrm{cov}(x, x') = \langle [\widehat{g}(x) - \langle \widehat{g}(x) \rangle] [\widehat{g}(x') - \langle \widehat{g}(x') \rangle] \rangle. \qquad (11)$$

Of course, once the covariance has been computed, the variance at any point x is given by $\mathrm{var}(x) = \mathrm{cov}(x, x)$. Evaluation of Eq. 11 is straightforward as $\widehat{g}(x)$ is given by

$$\widehat{g}(x) = \sum_{n=0}^{N-1} (\langle g(x_n) \rangle + n(x_n)) \sigma_N(x - x_n), \qquad (12)$$

where $\sigma_N(x)$ is given by Eq. 2, and thus

$$\mathrm{cov}(x, x') = \left\langle \left[\sum_{n=0}^{N-1} n(x_n)\sigma_N(x - x_n)\right] \times \left[\sum_{m=0}^{N-1} n(x_m)\sigma_N(x' - x_m)\right] \right\rangle, \qquad (13)$$

which can be rewritten

$$\mathrm{cov}(x, x') = \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} [\sigma_N(x - x_n)\sigma_N(x' - x_m) \times \langle n(x_n)n(x_m) \rangle]. \qquad (14)$$

Without additional assumptions about the noise, this expression cannot be simplified any further. In Section III we will consider simplifications to this expression for white noise and uncorrelated Poisson noise.

For periodic spline interpolation, the calculation of a covariance function is complicated by the fact that the points x and x' at which the covariance is being evaluated will in general fall into different intervals of the spline and the spline's value at these points will thus be specified by functions corresponding to different elements of the vector $\widehat{\mathbf{g}}(x)$. This is best handled by thinking in terms of computing an NxN matrix of covariance functions whose elements correspond to the possible combinations of pairs of intervals into which x and x' could fall. That is, given points x and x' between which it is desired to evaluate the covariance, one would compute the two intervals i and j in which these points fall and then evaluate the covariance function found in the ij^{th} position of the matrix. This matrix, which we denote with capital letters $\text{COV}(x, x')$, is given by

$$\text{COV}(x, x') = \left\langle [\widehat{\mathbf{g}}(x) - \langle\widehat{\mathbf{g}}(x)\rangle][\widehat{\mathbf{g}}(x') - \langle\widehat{\mathbf{g}}(x')\rangle]^T \right\rangle, \quad (15)$$

or using Eq. 9

$$\text{COV}(x, x') = \Big\langle \big[A + \mathcal{D}(x)B + \mathcal{D}^2(x)C + \mathcal{D}^3(x)D\big]$$
$$\times \mathbf{nn}^T \big[A + \mathcal{D}(x')B + \mathcal{D}^2(x')C + \mathcal{D}^3(x')D\big]^T \Big\rangle, \quad (16)$$

where $\mathbf{n}$ is the N-element vector with entries $n(x_n)$, which of course reduces to

$$\text{COV}(x, x') = \big[A + \mathcal{D}(x)B + \mathcal{D}^2(x)C + \mathcal{D}^3(x)D\big]$$
$$\times \langle\mathbf{nn}^T\rangle \big[A + \mathcal{D}(x')B + \mathcal{D}^2(x')C + \mathcal{D}^3(x')D\big]^T. \quad (17)$$

Again, without further assumptions about the noise, this expression cannot be simplified any further. In Section III we will consider simplifications to this expression for white noise and uncorrelated Poisson noise.

B. Monte Carlo Simulations

To confirm the analytic results derived above and applied below to the cases of white and uncorrelated Poisson noise, we performed Monte Carlo simulations employing each of these kinds of noise. We first calculated analytically 128 equispaced samples of a typical periodic function encountered in tomography: the angular function corresponding to one bin of the sinogram of a Shepp-Logan head phantom [15]. This function is shown in Fig. 1. We generated R=50,000 noise realizations of these samples contaminated with additive, zero-mean Gaussian noise ($\sigma_o^2 = 16$) and 50,000 others with Poisson noise. We then interpolated each of these realizations to 1024 samples using the methods discussed above. For CST and ZP interpolation, we took $K = 63$, the largest value it could have while still being below the Nyquist frequency of the 128 samples. We calculated the sample variance $\widehat{\sigma}^2(x_j)$ at each interpolated point x_j using

$$\widehat{\sigma}^2(x_j) = \frac{1}{(R-1)} \sum_{i=1}^{R} \left\{ \widehat{g}_{(i)}(x_j) - \frac{1}{R}\sum_{i=1}^{R} \widehat{g}_{(i)}(x_j) \right\}^2, \quad (18)$$
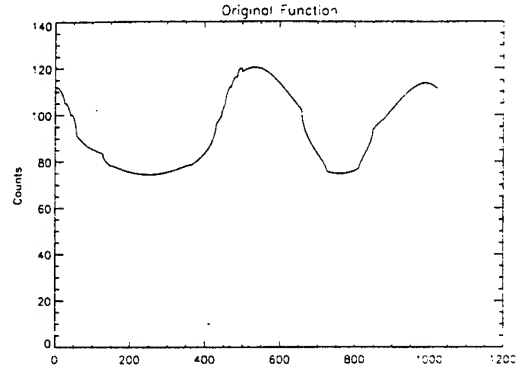

Figure 1: This periodic function, which is the angular function corresponding to one bin of a Shepp-Logan head phantom sinogram, was sampled 128 times and the samples then contaminated by Gaussian or Poisson noise prior to using CST/ZP and periodic spline interpolation to interpolate a sequence 8 times as dense.

where $\widehat{g}_{(i)}(x)$ is the curve interpolated from the ith noise realization. Similarly, we computed the sample covariance between each of 9 samples and the 1024 interpolated samples. This sample covariance between two points x_j and x_k was computed using

$$\widehat{C}(x_j, x_k) = \frac{1}{R-1} \sum_{i=1}^{R} \left\{ \widehat{g}_{(i)}(x_j) - \frac{1}{R}\sum_{i=1}^{R} \widehat{g}_{(i)}(x_j) \right\}$$
$$\times \left\{ \widehat{g}_{(i)}(x_k) - \frac{1}{R}\sum_{i=1}^{R} \widehat{g}_{(i)}(x_k) \right\}. \quad (19)$$

C. Image Noise Power Spectra

As mentioned in the introduction, periodic interpolation methods can be applied to the problem of interpolating additional projection views in few-view tomography. While the noise properties of the interpolated sinogram can be inferred from the analytic derivations discussed above, it is just as important to consider the noise properties of the images reconstructed from such sinograms, because these noise properties, as characterized by a noise power spectrum (NPS), greatly influence the detectability of small objects [16]. For this reason, we wished to calculate the NPS of images reconstructed after sinogram interpolation by CST/ZP and PS interpolation, as well as the NPS of images reconstructed directly from both a small and large number of views.

Though the noise power spectrum is strictly defined only for stationary noise processes, and the noise in images reconstructed from interpolated sinograms is not stationary due to the correlation between projections, there is precedent for examining the so-called average power spectrum of nonstationary processes [17]. To estimate the power spectrum for each approach, we generated 1000 sinograms containing only white, Gaussian noise ($\sigma_o^2 = 10$) and consisting of 128 projections bins and either 30 or 120 projection angles depending on which reconstruction approach was to be used. The sinograms were then either reconstructed directly by FBP or interpolated to 120 views using CST/ZP and PS

interpolation and then reconstructed by FBP. Each resulting noise image was multiplied by a circularly symmetric cosine window approximately 30 pixels in diameter and a 2D FFT computed. The resulting spectra were averaged and scaled so that the the volume under the power spectrum equaled the average variance in a circle of diameter 30 pixels.

III. RESULTS

A. White Noise

For CST and ZP interpolation, if the noise in the measured samples is assumed to be white, with constant variance σ_o^2 at every point, then in Eq. 14, $\langle n(x_n)n(x_m)\rangle = \sigma_o^2 \delta_{nm}$, where δ_{nm} is the Kronecker delta function and we may rewrite Eq. 14 as

$$\text{cov}(x, x') = \sigma_o^2 \sum_{n=0}^{N-1} \sigma_N(x - x_n)\sigma_N(x' - x_n). \quad (20)$$

This expression may be further simplified by comparing the sum with Eq. 1 and realizing that it can be viewed as a CST interpolation of the function $\sigma_N(x - x')$ sampled at points $x' = x_n$. Because $\sigma_N(x - x')$ is periodic and bandlimited to a frequency $K < N/2$, the interpolation is exact, and thus we conclude that .

$$\text{cov}(x, x') = \sigma_o^2 \sigma_N(x - x'). \quad (21)$$

The covariance is thus seen to be shift invariant, depending only on the difference $x - x'$ between the positions of any two points in the interpolated curve and not on their absolute positions. From this result, the variance is easily obtained:

$$\text{var}(x) = \text{cov}(x, x) = \sigma_o^2 \sigma_N(0). \quad (22)$$

From the definition of $\sigma_N(x)$ in Eq. 2, we see that $\sigma_N(0) = (2K+1)/N$ and thus conclude that when applying CST and ZP interpolation to samples corrupted by white noise, the variance of the interpolated curve is constant everywhere and equal to $(2K+1)/N$ times the variance in the original samples. Because $N \geq 2K+1$, this factor is less than or equal to 1, with equality when N minimally satisfies the Nyquist condition.

The Monte Carlo studies support the analytic findings of Eqs. 21 and 22. Figure 2(a) depicts a portion of the 1024-point sample variance curve for CST and ZP interpolation. For comparison, the analytic prediction is shown in Fig. 2(b). The sample variance is seen to be approximately constant at all points, in agreement with the analytic prediction, and it is found to have average value 15.933, in agreement with the analytic prediction of 15.875. Figures 2(c) and 2(d) depict the sample and predicted covariance relative to a fixed point located midway between two measured samples. While they are not shown, the sample covariance curves relative to fixed points at other positions in the interval between measured points were found to have essentially identical shapes, as expected from the shift-invariant form of the analytic prediction.

For spline interpolation, the assumption of white noise with variance σ_o^2 implies that in Eq. 17, the outer product
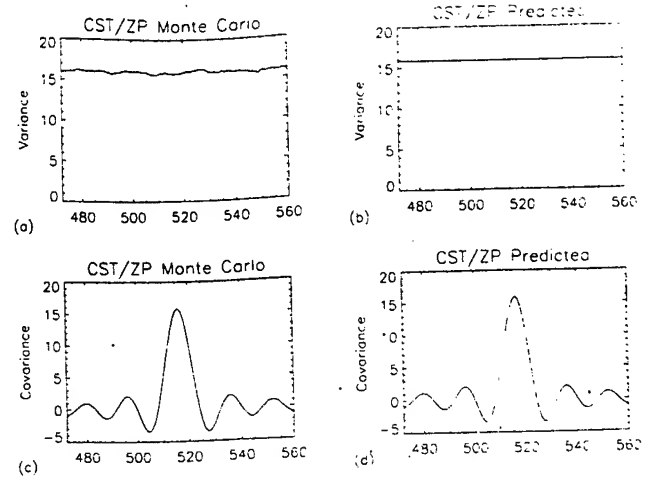

Figure 2: (a) Portion of sample variance curve for 50,000 realizations of CST/ZP interpolation from 128 samples to 1024 samples in the presence of additive, zero-mean Gaussian noise with $\sigma_o^2 = 16$. (b) Portion of analytically predicted variance curve for this task. (c) Portion of sample covariance curve relative to a fixed point midway between two measured points. (The shape of the curve for fixed points at other positions in the interval between measured samples was identical.) (d) Portion of analytically predicted covariance curve.
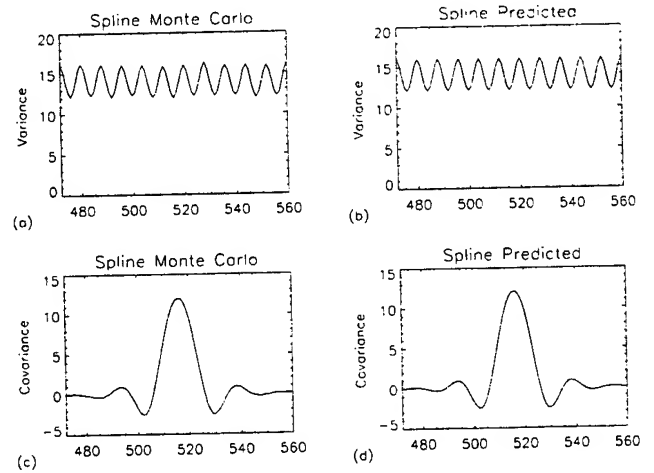

Figure 3: (a) Portion of sample variance curve for 50,000 realizations of periodic spline interpolation from 128 samples to 1024 samples in the presence of additive, zero-mean Gaussian noise with $\sigma_o^2 = 16$. (b) Portion of analytically predicted variance curve for this task. The maxima in these curves occur at the locations of the measured samples. (c) Portion of sample covariance curve relative to a fixed point midway between two measured points. (The shape of the curve for fixed points at other positions in the interval between measured samples was quite different.) (d) Portion of analytically predicted covariance curve for a fixed point midway between two measured points.

$\langle \mathbf{n}^T\mathbf{n}\rangle = \sigma_o^2 I$, where I is the identity matrix. The properties of the covariance expression given in Eq. 17 cannot be discerned by inspection and the result must be evaluated numerically. In particular, to obtain the variance of the interpolated function, the diagonal elements of the covariance matrix must be evaluated at $x = x'$. Figure 3(b) illustrates the result of such an evaluation. It is clearly seen that unlike CST and ZP

interpolation, the variance is not constant at interpolated points, but rather falls smoothly to a minimum at points midway between the original measured samples, while remaining unchanged at the measured points. The Monte Carlo results shown in Fig. 3(a) conform quite closely to this prediction. Figures 3(c) and 3(d) show a portion of the covariance function relative to a fixed point located midway between two measured samples. It is seen to have a slightly wider central lobe than the equivalent curve for CST/ZP interpolation but also to die out more quickly beyond this central lobe. Unlike in the CST/ZP case, the covariance curves relative to fixed points at other positions in the the interval between measured points were found to have quite different shapes, becoming asymmetric as the fixed point moves away from the center of the interval. Thus, the covariance of the interpolated curve is not shift-invariant for spline interpolation.

B. Poisson Noise

For CST and ZP interpolation, if the noise in the measured samples is assumed to be uncorrelated Poisson, then in Eq. 14, $\langle n(x_n) n(x_m) \rangle = \langle g(x_n) \rangle \delta_{nm}$, and we may rewrite Eq. 14 as

$$\text{cov}(x, x') = \sum_{n=0}^{N-1} \langle g(x_n) \rangle \sigma_N(x - x_n) \sigma_N(x' - x_n). \quad (23)$$

This sum can be viewed as a CST interpolation, this time of the function $\langle g(x') \rangle \sigma_N(x - x')$. However, this function, while periodic, is not in general bandlimited. However, if $\langle g(x') \rangle$ is slowly varying, at least over the interval between x and x', it can be assumed that the product is approximately bandlimited and thus that

$$\text{cov}(x, x') \cong \langle g(x') \rangle \sigma_N(x - x'). \quad (24)$$

The covariance function is seen to be approximately shift invariant, certainly as the fixed point is moved through any single interval between measured samples. The variance is again easily obtained:

$$\text{var}(x) = \text{cov}(x, x) \cong \langle g(x) \rangle \sigma_N(0). \quad (25)$$

Because $\sigma_N(0) = (2K + 1)/N$, we find that the variance in the interpolated curve is flat locally (again, certainly between the positions of any two measured samples), while globally it tracks the variance in the measured samples.

The Monte Carlo studies confirm the analytic predictions of Eqs. 24 and 25. Figures 4(a) and 4(b) illustrate how the variance in the interpolated function, while locally constant, follows the variance in the measured samples over larger distances. The covariance function (not shown) is found to behave over short distances much as it did in the white noise case.

For spline interpolation of data contaminated by Poisson noise, the same general trends as in the white noise case are observed. The variance in the interpolated curve again falls smoothly to a minimum between measured samples, though now it tracks the variance in the measured samples globally.
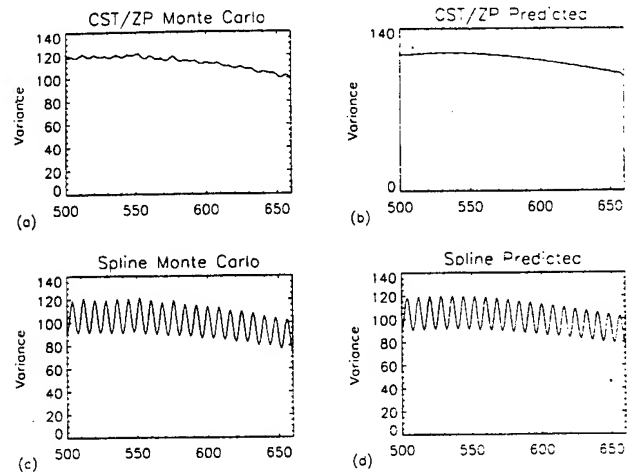

Figure 4: (a) Portion of a sample variance curve for 50,000 realizations of CST/ZP interpolation from 128 samples to 1024 samples in the presence of Poisson noise. (b) Portion of analytically predicted variance curve for this task. (c) Portion of a sample variance curve for 50,000 realizations of periodic spline interpolation from 128 samples to 1024 samples in the presence of Poisson noise. (d) Portion of analytically predicted variance curve for this task.
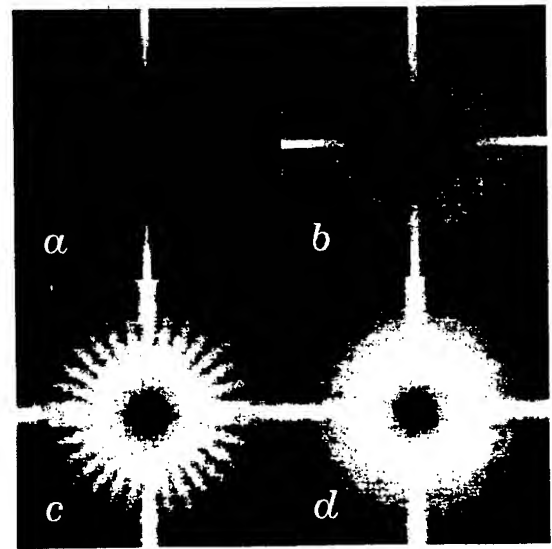

Figure 5: Power spectra for FBP reconstructions of stationary white Gaussian noise from (a) 30 projections. (b) 120 projections. (c) 120 projections interpolated by periodic spline interpolation from 30 projections. (d) 120 projections interpolated by zero-padding interpolation from 30 projections.

This is clearly seen in Figs. 4(c) and 4(d). The covariance curves were found to behave locally as they did in white noise case, changing form as the fixed point is swept through an interval between measured points.

C. Image noise power spectra

The spectra are shown in Fig. 5. The high-intensity streaks on the horizontal and vertical midlines are related to the use of linear interpolation in the radial direction during backprojection. The NPS for a direct reconstruction from 30 views shown in Fig. 5(a) is seen to consist of distinct rays

passing through the origin, as predicted by theory [18]. The NPS for a direct reconstruction from 120 views shown in Fig. 5(b) has a more continuous and circularly symmetric appearance. The NPS for reconstruction from 120 views interpolated from 30 views by PS interpolation (Fig. 5(c)) is seen to be quite non-uniform, a hybrid of the discrete spokes of the 30-angle reconstruction and the more uniform appearance of the 120-view reconstruction. The NPS for reconstruction from 120 views interpolated from 30 views by CST/ZP interpolation (Fig. 5(d)), on the other hand, is seen to resemble the 120-view NPS quite closely, being esentially continuous and circularly symmetric.

IV. DISCUSSION AND CONCLUSIONS

The differences between the CST/ZP and periodic spline interpolation approaches emerge most clearly in the white noise case discussed above. Here we saw that CST/ZP interpolation results in a curve with constant variance at all points (a factor of $(2K + 1)/N$ times the variance in the original measured samples), and having a covariance function that depends only on the distance $x - x'$ between two points. Put simply, if the noise in the measured samples is uncorrelated and wide-sense stationary, then the noise in the interpolated curve is also wide-sense stationary. Naturally, the noise does not remain uncorrelated. The situation is quite different for periodic spline interpolation. Here we saw that the variance remains equal to the variance in the original samples only at the interpolated points corresponding to those samples, and that it falls to a minimum at the midpoint between two such samples. Moreover, it was observed that the covariance function does not simply depend on the distance $x - x'$ between two points but depends also on the positions of the two points within their respective intervals between measured samples. Thus though the noise in the measured samples may be uncorrelated and wide-sense stationary, the noise in the curve interpolated by periodic spline interpolation is neither.

It is well known that when a stationary process is input to a linear, shift-invariant system, the output process is stationary as well [19]. Both CST/ZP and PS interpolation can be shown to involve linear, shift-invariant operations on the data, yet we have seen that given uncorrelated, stationary white noise in the measured samples, CST/ZP interpolation yields a stationary curve and PS interpolation does not. How do we explain this apparent paradox? The answer lies in the fact that the set of measured samples represents a discrete stationary process while the result of interpolation is a continuous process, and one must take care in comparing them. Systems theory makes no claim about the ability of a linear, shift-invariant system to turn a discrete stationary process into a continous one. Indeed, if the original discrete process is viewed in the continuous domain it is not at all stationary: the variance falls to zero in the gaps between measured samples. What is truly surprising, then, is not that periodic spline interpolation fails to produce a stationary continuous process from the stationary discrete samples, but rather that CST/ZP interpolation succeeds in doing so.

The decision to use one approach or the other for the interpolation of additional angular views in few-view tomography should take into consideration the particular requirements of the study being performed. Given that the interpolatory accuracy of the two approaches is generally comparable, at first glance it would seem that the variance-reducing properties of the periodic spline interpolation might be preferable. However, this comes at the cost of having widely varying variance levels in neighboring projections, as well as non shift-invariant covariance. These non-uniformities do not noticeably affect the resulting image quality if one reconstructs directly from the interpolated sinogram and thus periodic spline interpolation may be appropriate if the images are only to be inspected visually, and particularly if the task involves the examination of relatively large structures. If, however, one seeks to perform any sort of principled smoothing on the projections or quantitation on the reconstructed image that requires knowledge of the statistical properties of the projections, then the stationarity-preserving properties of CST/ZP interpolation may be the best choice.

CST/ZP interpolation may also be the best choice if the task involves the detection of small objects. As mentioned in Section II, the noise properties of reconstructed images, as characterized by a noise power spectrum (NPS), greatly influence the detectability of small objects [16]. We examined the NPS of images reconstructed after sinogram interpolation by CST/ZP and PS and found the CST/ZP NPS to be much more similar to the NPS of images reconstructed from a full complement of 120 original views than was the spline NPS. Indeed, the spline NPS demonstrated the existence of non-uniform angular correlations that could potentially hinder detection tasks.

Finally, the ability of CST and ZP interpolation to reduce the variance in the interpolated curve by the factor $(2K + 1)/N$ relative to the variance in the measured samples deserves comment. This is simply an implicit exploitation of the ability to achieve noise reduction through oversampling and filtering [20, Ch. 5], and the mechanism is easiest to appreciate in the case of zero-padding interpolation for stationary white noise. The DFT of Eq. 3 has N terms, corresponding to frequencies out to $k = \pm N/2$ for N even or $k = \pm(N - 1)/2$ for N odd. In Eq. 4, however, we see that all DFT components corresponding to frequencies $|k| > K$ are explicitly set equal to zero, because $g(x)$ is assumed to be bandlimited (at least approximately) to frequency K. However, if the measured samples are corrupted by white noise, then their DFT can be seen as the sum of the true DFT and the DFT of the noise process, which has contributions at all N frequency components. Zeroing out all but $2K + 1$ of these would thus be expected to reduce the noise magintude by the factor $(2K + 1)/N$.

V. ACKNOWLEDGMENT

VI. References

[1] R. A. Brooks, G. H. Weiss, and A. J. Talbert, "A new approach to interpolation in computed tomography," *J. Comput. Assist. Tomogr.*, vol. 2, pp. 577–585, 1978.

[2] H. Stark, J. W. Woods, I. Paul, and R. Hingorani, "Direct Fourier reconstruction in computer tomography," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 29, pp. 237–245, 1981.

[3] H. Stark, "Sampling theorems in polar coordinates," *J. Opt. Soc. Am.*, vol. 69, pp. 1519–1525, 1979.

[4] H. Stark and M. Wengrovitz, "Comments and corrections on the use of polar sampling theorems in CT," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 31, pp. 1329–1331, 1983.

[5] R. W. Schafer and L. R. Rabiner, "A digital signal processing approach to interpolation," *Proc. IEEE*, vol. 61, pp. 692–702, 1973.

[6] K. P. Prasad and P. Satyanarayana, "Fast interpolation algorithm using FFT," *Electron. Lett.*, vol. 22, pp. 185–187, 1986.

[7] D. Fraser, "Interpolation by the FFT revisited—an experimental investigation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 665–675, 1989.

[8] T. J. Cavicchi, "DFT time-domain interpolation," *IEE Proceedings-F*, vol. 139, pp. 207–211, 1992.

[9] P. J. La Rivière and X. Pan, "Mathematical equivalence of zero-padding and circular sampling thorem interpolation with implications for direct Fourier image reconstruction," in *Proc. SPIE*, vol. 3338, pp. 1117–1126, 1998.

[10] H. Spath, *One-Dimensional Spline Interpolation Algorithms*. Wellesley, MA: A. K. Peters, 1993.

[11] C. H. Reinsch, "Smoothing by spline functions," *Numer. Math.*, vol. 10, pp. 177–183, 1967.

[12] P. J. La Rivière and X. Pan, "Comparison of angular interpolation approaches in few-view tomography using statistical hypothesis testing," in *Proc. SPIE*, vol. 3661, 1999. (To appear).

[13] E. O. Brigham, *The Fast Fourier Transform and Its Applications*. Englewood Cliffs: Prentice Hall, 1988.

[14] P. J. Green and B. W. Silverman, *Nonparametric Regression and Generalized Linear Models*. London: Chapman Hall, 1994.

[15] L. A. Shepp and B. F. Logan, "The Fourier reconstruction of a head section," *IEEE Trans. Nucl. Sci.*, vol. 21, pp. 21–43, 1974.

[16] International Comission on Radiation Units and Measurements, "Medical imaging: The assessment of image quality," 7910 Woodmont Ave., Bethesda, MD, 20814, 1996.

[17] S. C. Moore, M. F. Kijewski, S. P. Müller, and B. L. Holman, "SPECT image noise power: Effects of nonstationary projection noise and attenuation compensation," *J. Nucl. Med.*, vol. 29, pp. 1704–1709, 1988.

[18] K. Hanson, "Detectability in computed tomographic images," *Med. Phys.*, vol. 6, pp. 441–451, 1979.

[19] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. New York: McGraw Hill, 1991.

[20] R. J. Marks, *Introduction to Shannon Sampling and Interpolation Theory*. New York: Springer Verlag, 1991.

Few-View Tomography Using Roughness-Penalized Nonparametric Regression and Periodic Spline Interpolation

P. J. La Rivière, *Student Member, IEEE,* and X. Pan, *Member, IEEE*

Department of Radiology, The University of Chicago, Chicago, IL, 60637

Abstract

The ability to reconstruct high-quality tomographic images from a smaller number of projections than is usually used could reduce imaging time for many nuclear-medicine studies. This would particularly benefit studies such as cardiac SPECT where patient motion during long acquisitions can lead to motion artifacts in the reconstructed images. To this end, we have investigated sinogram pre-processing techniques designed to enable filtered backprojection (FBP) to produce high-quality reconstructions from a small number of views. Each projection is first smoothed by performing roughness-penalized nonparametric regression using a generalized linear model that explicitly accounts for the Poisson statistics of the data. The resulting fit curves are natural cubic splines. After smoothing, additional angular views are generated using periodic spline interpolation, and images are reconstructed using FBP. The algorithm was tested on data from SPECT studies of a cardiac phantom placed at various radial offsets to enable examination of the algorithm's dependence on the radial extent of the object being imaged.

I. INTRODUCTION

In routine nuclear-medicine tomographic studies, there is usually a tradeoff between image quality and imaging time. Increasing the number of angular views acquired, the number of counts per view, or both will generally improve image quality but will also lengthen imaging time. In general, concerns about image quality take precedence over concerns about imaging time. However, in studies where the patient must assume an awkward or uncomfortable position, long acquisition times can potentially lead to patient motion and thus to motion artifacts in the reconstructed images. One way to reduce imaging time without significant sacrifice of image quality is to use a continuous acquisition mode, in which the time wasted in moving the camera between views is eliminated. However, continuous acquisition is not recommended for some of the studies most plagued by motion artifacts, such as gated cardiac SPECT. In such studies, reducing imaging time must be accomplished either by reducing the number of angular views or by reducing the number of counts per view. In this paper, we focus on the first approach, investigating algorithms tailored to generate diagnostically useful images from a smaller number of angular views than is usually used while holding the number of counts per view constant.

One strategy for reconstructing acceptable images in few-view tomography is to incorporate into the reconstruction process as much prior information as possible about the expected image. For instance, constraints regarding the size [1], symmetry properties [2], or even the mean [3] of the expected image may be used. However, this approach generally leads to iterative algorithms involving multiple reconstruction and reprojection steps that are computationally intensive and subject to concerns regarding convergence and regularization. Moreover, the incorporation of prior information may involve a subjective and time-consuming operator-dependent step. These drawbacks are generally well justified by the resulting image quality in extreme situations involving reconstruction of a fairly complex object from 10 or fewer projections, where it is hardly possible to generate acceptable images without incorporating prior information. However, our interest is in more moderate situations, such as reconstruction from 60 or 30 views when 120 might normally be used, and in these situations, a more computationally efficient, fully automatic, robust algorithm may be preferable. For these reasons, we focus in this work on the development of algorithms for few-view reconstruction involving sinogram pre-processing followed by reconstruction by filtered backprojection (FBP).

In the absence of additional *a priori* constraints about the object being imaged, the minimum number of angular views required to produce an accurate tomographic reconstruction of a given object using FBP is dictated by two factors. First, the angular sampling of the object's sinogram must satisfy the Nyquist sampling condition. Absent this, any reconstruction is doomed to suffer from angular aliasing artifacts. Second, the number of angular samples must satisfy FBP's implicit assumptions about the density of angular sampling. It is well known that FBP reconstructions from a small number of angular views are degraded by prominent star-shaped artifacts. These two conditions are not in general equivalent, as can be appreciated most keenly when considering the case of imaging a circularly symmetric object. In this instance, a single projection view is sufficient to satisfy the Nyquist sampling condition, while a FBP reconstruction from this single view would be an uninterpretable set of parallel streaks. Indeed, Brooks *et al.* have shown that for a circularly symmetric object imaged with N bins per projection, a minimum of $\sim 1.1\pi N/4$ projections must be acquired over 180° to produce an essentially artifact-free reconstruction using FBP [4].

The Nyquist condition is the more fundamental of the two sampling requirements discussed above, because when it is satisfied by a number of samples less than the number required by the reconstruction algorithm, it is in principle possible to interpolate exactly the additional views needed. This is only strictly true in the absence of noise, of course, a condition that rarely obtains in emission tomography. In the presence of noise, angular interpolation usually leads to severe circular artifacts, because the noise process is rarely if ever bandlimited to the Nyquist frequency of the samples.

However, by smoothing each projection prior to interpolation, the noise-engendered interpolation artifacts can be effectively eliminated. While simple linear, shift-invariant smoothing filters, such as a Hanning window, lead to reasonable results, we opted to explore a more sophisticated smoothing approach that makes better use of the information contained in the data. The approach we chose was roughness-penalized nonparametric regression using an explicit Poisson statistical model [5], which leads to fit functions that are natural cubic splines, piecewise cubic polynomials that are continuous up to and including the second derivative and satisfying the so-called natural boundary conditions. While this approach is similar in spirit to the information-weighted splines of Fessler [6], the use of the explicit Poisson model in the present case leads to a rather different, iterative algorithm. Moreover, rather than choosing the smoothing parameter *a priori,* we have implemented an automatic algorithm for determining it, based on the principle of cross-validation and adapted appropriately for Poisson-distributed data. The subsequent interpolation of additional views is also performed using splines, this time satisfying periodic boundary conditions. Image reconstruction then proceeds as usual, with due recognition of the fact that the sinogram has already been smoothed.

II. METHODS

A. Interpolation of angular views

Consider a tomographic acquisition that yields a two-dimensional discrete sinogram $p(\xi_n, \phi_m)$, where ξ_n, $n = 1, \ldots, N$, is the projection bin and ϕ_m, $m = 1, \ldots, M$, is the projection angle. The number M of angular samples is assumed to satisfy the Nyquist condition at least approximately (i.e., the energy of the angular spectrum beyond the Nyquist frequency is assumed to fall below some small threshold). We wish to increase the number of angular samples to KM, where K is an integer and KM is large enough to eliminate star artifacts from an FBP reconstruction, by interpolating additional views between the measured ones. To do this, the sinogram is viewed as a set of 1D sampled functions of projection angle, each labeled by a projection bin ξ_n, with the samples denoted by $p_{\xi_n}(\phi_m)$. Then continuous one-dimensional (1D) interpolating curves $p_{\xi_n}^i(\phi)$, where the superscript i indicates interpolated, are fit to each of these sampled functions and resampled to obtain the additional angular views.

Ideally, a periodic interpolation method should be chosen in order to make use of the inherent periodicity of the angular samples. In image reconstruction, periodic interpolation has primarily been studied in the context of direct Fourier reconstruction, where there is a need to interpolate from a polar to a Cartesian grid in the Fourier space of the image function while exploiting the periodicity of the azimuthal samples [7–9]. Fewer studies exist of interpolation in sinogram space [1, 10], though many of the Fourier-space interpolation techniques can be readily applied in sinogram space, including simple schemes such as nearest-neighbor

and linear interpolation with periodic boundary conditions as well as the more complex circular sampling theorem (CST) and zero-padding approaches [11–14]. Despite its simplicity, zero-padding interpolation is quite accurate; indeed, it can be shown [15] that it is mathematically equivalent to CST interpolation and that both are exact when interpolating bandlimited, periodic functions whose samples satisfy the Nyquist condition. Because ZP interpolation exploits the efficient FFT algorithm, it has a much lower computational burden than CST interpolation, and is thus to be preferred in this context.

Despite the theoretical exactness of CST/ZP interpolation when the Nyquist condition is satisfied, the two methods are quite sensitive to mild violations of the criterion, giving rise to the well-known Gibbs oscillations that radiate widely from sharp edges [16]. This shortcoming prompted us to explore yet another interpolation technique based on cubic splines that while not theoretically exact is known to be very accurate and quite robust in the face of violations of the Nyquist condition. Our confidence in the use of splines was bolstered by studies we performed [17] of 1D interpolation in which samples of known analytic functions representing the angular variations of projections of Shepp-Logan-like and breast phantoms [18] were interpolated at intermediate points and the interpolated values then compared to the known values at those points. In these tests, spline interpolation was found to have a statistically significant lower root-mean-squared error than linear interpolation or CST/ZP interpolation for virtually all sampling intervals.

A periodic cubic spline is a curve comprised of generally different third-order polynomials between each pair of known abscissas ϕ_m and ϕ_{m+1}, with the overall curve being continuous up to and including the second derivative at each abscissa [19]. Naturally, in the case of an interpolating spline (as opposed to a smoothing spline), the curve $p_{\xi_n}^i(\phi)$ is also constrained to pass through the known ordinate values $p_{\xi_n}(\phi_m)$ at each abscissa ϕ_m. The spline can be represented as

$$p_{\xi_n}^i(\phi) = a_m + b_m\phi + c_m\phi^2/2 + d_m\phi^3/3, \qquad (1)$$

for $\phi \in [\phi_m, \phi_{m+1}]$, where $m = 1, \ldots, M$. The process of fitting an interpolating spline is then tantamount to solving a set of linear equations for the defining coefficients a_m, b_m, c_m, and d_m in each interval $[\phi_m, \phi_{m+1}]$ such that the continuity and interpolation conditions are satisfied. For a periodic spline, the coefficients must also satisfy periodic boundary conditions.

In order to illustrate the effects of this interpolation and to motivate our approach to mitigating noise, we have reconstructed images from simulated projections of a numerical phantom. The true phantom is shown in Fig. 1(a). In Fig. 1(b), the phantom is shown reconstructed by ramp-filtered FBP from 120 noiseless angular views. In Fig. 1(c), a reconstruction from 30 views, we observe the star-shaped artifacts discussed in the introduction. Figure 1(d) illustrates the results of ramp-filtered FBP reconstruction of the phantom after interpolating from 30 to 120 angular views using periodic spline interpolation. We see that the star-shaped artifacts have
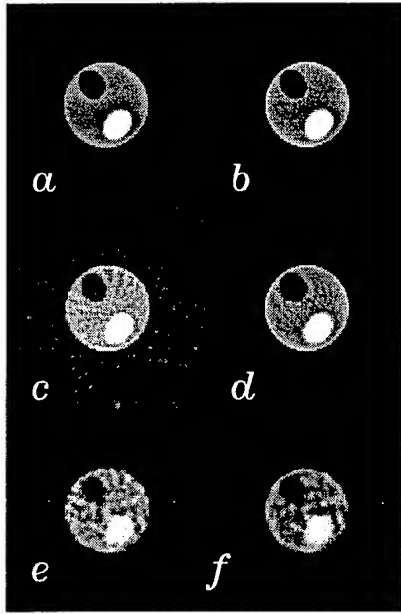
Figure 1: Reconstructions of a numerical phantom. (a) True phantom. (b) Reconstruction by ramp-filtered FBP from 120 noiseless angular views. (c) Reconstruction by ramp-filtered FBP from 30 noiseless angular views. (d) Reconstruction by ramp-filtered FBP after interpolation from 30 to 120 noiseless angular views. (e) Reconstruction by Hanning-filtered FBP from 30 angular views with Poisson noise. (f) Reconstruction by Hanning-filtered FBP after interpolation from 30 angular views with Poisson noise to 120 views.

been essentially eliminated. In Fig. 1(e), we show the results of a Hanning-filtered reconstruction from 30 angular views with Poisson noise (125,000 total counts). The star-shaped background artifacts are still present and we observe the image to be somewhat noisy. Finally, in Fig. 1(f), we show the results of a Hanning-filtered reconstruction of the phantom after interpolating from 30 to 120 views. We observe that the star-shaped artifacts have been eliminated as before, but that circular noise artifacts have appeared in the image.

B. Smoothing

While the results of spline interpolation in the noiseless case (Fig. 1(d)) are encouraging, the poor reconstruction in the noisy case (Fig. 1(f)) indicates that prior smoothing of the sinogram may be necessary if spline interpolation of angular views is to succeed in the presence of noise. Rather than simply apply a linear, shift-invariant filter to the noisy projection data, we decided to explore a more principled statistical approach using roughness-penalized nonparametric regression based on a generalized linear model (GLM) [5] that explicitly accounts for the Poisson nature of the SPECT data.

1) Nonparametric regression using a GLM

Regression analysis with a single explanatory variable considers the problem of fitting a curve to a set of data pairs (Y_i, x_i), $(i = 1, \ldots, N)$, where the Y_i are the measured values of the quantity of interest and the x_i the corresponding values of the explanatory variable. The variation in the Y_i is assumed to have two components: a systematic component captured by a vector of predictors θ_i that depends on the x_i, and a

random component specifying the distribution of the Y_i given θ_i. In classical linear regression, for example, the systematic component is assumed to be of the form $\theta_i = ax_i + b$, and the Y_i are assumed to be normally distributed about the θ_i, i.e., $Y_i \sim N(\theta_i, \sigma^2)$. In general, maximum-likelihood methods are used to estimate regression curves, and in the case of classical linear regression, the maximum-likelihood estimates of a and b are easily shown to be those that minimize the sum of squares $\sum_{i=1}^{N}(Y_i - ax_i - b)^2$ [20].

Nonparametric regression using a GLM relaxes both of the assumptions of classical linear regression. First, it eliminates the assumption that the predictors θ_i depend on the explanatory variable in a simple parametric way, representing them instead as an arbitrary function of the explanatory variables: $\theta_i = g(x_i)$. This would seem to complicate the problem immensely, turning a relatively straightforward finite-dimensional estimation problem into an intractable infinite-dimensional one. However, in practice the problem is made tractable by adding the further constraint that the estimated curve $g(x)$ be smooth and enforcing this constraint by penalizing the likelihood with a term of the form $\int g''(x)^2 dx$. If the unpenalized likelihood depends on $g(x)$ only through its values $g(x_i)$ at the measured points x_i, $i = 1, \ldots, N$ (which is usually the case), it can be shown that the minimizer of the penalized likelihood is necessarily a natural cubic spline [5]. These are simply cubic splines that are constrained to be linear outside the set of measured points, and like the periodic splines discussed above, they can be specified by a finite number of coefficients. (In this sense, roughness-penalized nonparametric regression does yield a parametrized curve, but given that there are at least as many parameters as observations, the spirit is rather different from that of standard parametric techniques.)

Nonparametric regression using a GLM also relaxes the assumption that the data are normally distributed in favor of the much broader class of exponential distributions, which have probability densities of the form

$$p(y_i \mid \theta_i, \phi) = \exp\left(\frac{y_i\theta_i - b(\theta_i)}{\phi} + c(y_i, \phi)\right), \quad (2)$$

where θ_i is the so-called natural parameter of the exponential family and ϕ is a scale parameter [21]. This family comprises many well-known distributions, each corresponding to a different choice of the functions b and c. Of particular interest to emission tomography is the choice $b(\theta_i) = e^{\theta_i}$, $\phi = 1$, $c(y_i, \phi) = -\log(y_i!)$, which corresponds to a Poisson distribution with parameter $\lambda_i = e^{\theta_i}$. If a nonparametric dependence $\theta_i = g(x_i)$ of the predictor on the explanatory variables is assumed, the log-likelihood of N independent observations Y_i drawn from a Poisson density is given by

$$\ell(g, \phi) = \sum_{i=1}^{N} \left(Y_i g(x_i) - \exp\left[g(x_i)\right] - \log(Y_i!)\right). \quad (3)$$

The goal of roughness-penalized nonparametric regression is to estimate the curve $g(x)$ that maximizes this log-likelihood

subject to the penalty $\int g''(x)^2 dx$, i.e. to maximize

$$\sum_{i=1}^{N} \{Y_i g(x_i) - \exp[g(x_i)]\} - \frac{1}{2}\alpha \int g''(x)^2 dx, \quad (4)$$

where terms independent of g have been dropped and where α is the so-called smoothing parameter, to be discussed below. Note that $g(x)$ here represents an estimate of the *log* of the Poisson parameter $\lambda(x)$. As mentioned above, this expression can be shown to be maximized by a natural cubic spline and it can also be shown that for natural cubic splines, the penalty $-\frac{1}{2}\int g''(x)^2 dx = -\frac{1}{2}\alpha g^T K g$, where g is an N-element vector with $g_i = g(x_i)$, and K is an NxN matrix determined by the spacing of the measurement points x_i. Because the natural cubic spline interpolating any specified set of points $g(x_i)$ is unique, finding $g(x)$ is thus tantamount to finding g [5].

As is traditional in GLMs [21], we use Fisher scoring to find the g that maximizes Eq. 4, which yields the following iterative update equation

$$g^{(k+1)} = (W + \alpha K)^{-1} W z^{(k)}, \quad (5)$$

where z is an N-element vector with, for Poisson data, components

$$z_i^{(k)} = g_i^{(k)} + \frac{Y_i - \exp\left(g_i^{(k)}\right)}{\exp\left(g_i^{(k)}\right)}, \quad (6)$$

W is a diagonal matrix with, for Poisson data, entries $W_{ii} = \exp\left(g_i^{(k)}\right)$, and the superscript (k) refers to the kth iteration. The initial estimate $g^{(0)}$ is chosen to have components $g_i^{(0)} = \log\{\max(Y_i, \epsilon)\}$, where ϵ is a small positive constant introduced to avoid computing $\log(0)$. Iteration continues until the sum of the absolute changes in the components of g from one iteration to the next falls below a prespecified threshold. While this would seem to be a very computationally intensive procedure, the banded structure of some of the matrices involved can be exploited to keep the algorithm to $O(N)$.

2) Choice of the smoothing parameter

The choice of the smoothing parameter α profoundly influences the appearance of the fit curve $g(x)$, for α determines the relative influence of the two terms in the penalized-likelihood expression, the first rewarding goodness-of-fit to the data, the second rewarding smoothness. A small value of α leads to a ragged curve while a large value of α leads to a smooth curve. In most applications, a value between these two extremes is desirable, and while this can be found through trial and error for most datasets, a more principled and automatic approach would clearly be preferred.

One such automatic approach to choosing the smoothing parameter is based on the principle of cross validation (CV), which has been discussed in the context of image processing by Galatsanos and Katsaggelos [22]. The approach is grounded in

the assumption that the choice of α should yield a fit curve $g(x)$ that accurately predicts the outcomes of further observations. Remarkably, the predictive accuracy of the fit curve can be quantified solely on the basis of the fit values and the measured data [23]. This so-called CV score can be expressed as

$$CV(\alpha) = \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{Y_i - g(x_i)}{1 - A_{ii}(\alpha)}\right\}^2, \quad (7)$$

where A_{ii} are the diagonals of the hat matrix A, which links the values of the estimate at the measured points to the values of the observations at those points: $g = A(\alpha)Y$. The value of α minimizing this curve is considered optimal, and can generally be found fairly quickly using a golden section search minimization approach.

The CV score of Eq. 7, based on a residual sum of squares, is more appropriate for normally distributed data than for the more general class of distributions encompassed by GLMs. For GLMs, we follow O'Sullivan et al. [24], who propose replacing the residual sum of squares with the generalized Pearson χ^2 statistic, which for Poisson distributed data is given by

$$CV_{\text{GLM}}(\alpha) = \frac{1}{N}\sum_{i=1}^{N}\frac{(Y_i - \exp(g_i))^2 / \exp(g_i)}{(1 - A_{ii}(\alpha))^2}. \quad (8)$$

where the A_{ii} are the diagonal elements of a matrix A given by $(W + \alpha K)^{-1}W$, where W is as above, evaluated after the final iteration. Calculating the CV score would seem to be an $O(N^2)$ operation, for computing the matrix A involves inverting the matrix $(W + \alpha K)$. However, two facts work in our favor: first, $(W + \alpha K)$ can be manipulated to take advantage of band structures and second, only the diagonal elements of A are needed. Hutchinson and de Hoog [25] have developed an algorithm that allows the diagonal elements of the inverse of a band matrix to be computed in $O(N)$, and this is the approach we have used.

C. Data Acquisition and Processing

In order to examine the response of the algorithm to increasingly difficult interpolation tasks, we imaged a compact object placed at increasing radial offsets. This increases the bandwidth of the angular functions at each projection bin. Specifically, we acquired projections of a Data Spectrum ventricular phantom placed at five different radial offsets from the COR: 0, 5, 9, 12, and 15 cm. The phantom was filled with 121 MBq (3.27 mCi) of Tc-99m, contained a 1-cm defect insert, and was not placed within a water-filled torso phantom. We imaged this phantom with a Picker 3000XP three-headed SPECT system fit with low-energy, high-resolution, parallel-hole collimators, acquiring studies containing 120 angular views over 360°. We used a 25-cm radius circular orbit and step-and-shoot mode for all of the acquisitions; each head acquired to a 128x128 pixel image, though we preserved only the 32 slices spanning the phantom. A total of about 500,000 counts was collected. From this data we extracted 3D sinograms corresponding to 15, 30, 60, and 120 views, respectively. Thus we had 20 different sinograms,

corresponding to the 20 possible combinations of radial offset and number of angular views. We reconstructed images from these 20 sinograms using four different processing techniques:

1. No pre-smoothing of the sinogram and slice-by-slice reconstruction from available views by FBP using a Hanning filter (cutoff=0.8).

2. No pre-smoothing of the sinogram, spline interpolation from the available views to 120 angular views, and slice-by-slice reconstruction by FBP using a Hanning filter (cutoff=0.8).

3. Roughness-penalized nonparametric regression smoothing of the sinogram and slice-by-slice reconstruction from the available views by FBP using a ramp filter (cutoff=1.0).

4. Roughness-penalized nonparametric regression smoothing of the sinogram, spline interpolation from the available views to 120 views, and slice-by-slice reconstruction by FBP using a ramp filter (cutoff=1.0).

The particular application of the nonparametric regression technique to this data requires some explanation. The projection data for each of the radial offsets is a 3D sinogram of 128 bins, 32 slices, and 120 angles. While the smoothing technique was applied to each projection in each slice independently, we decided it would be wise to use the same smoothing parameter for all the projections. To do otherwise would have invited inconsistencies in the smoothed sinogram likely to produce streaks and other artifacts in the reconstructed images. At the same, we wished to select the smoothing parameter to be applied to the data using some form of cross-validation. The solution was to string together all of the 1D projections in the complete 3D sinogram into a single, long 1D function, and to find the smoothing parameter that minimized the CV score for that function when smoothed by the roughness-penalized nonparametric technique. This value of α was then used in smoothing each of the projections in this sinogram individually.

III. RESULTS

For ease of comparison and simplicity of presentation, we have grouped the reconstructed images by the number of angles in the original sinogram, and we show in Fig. 2 the results for only three of the radial offsets: 0, 9, and 15 cm. These three suffice to illustrate the trends observed. For each combination of number of angles and radial offset we show the results of reconstructing by use of each of the four techniques outlined above.

We observe that reconstructions from available views without pre-smoothing or interpolation display star-shaped artifacts and a mottled appearance when the number of views is small. Interpolation alone mitigates the star-shaped artifacts but leads to severe circular artifacts, particularly in the case of a small number of views and a large radial offset. Smoothing alone reduces the noise visibility but has little effect on the star-shaped artifacts. The combination of smoothing and interpolation still produces circular interpolation artifacts in the case of a large radial offset combined with a small number of original views, but these are less severe than when interpolation alone is used. Overall, though, visually appealing reconstructions result for less challenging combinations of radial offset and number of views, including as few as 15 angles in the 0-cm offset case.

While Fig. 2 demonstrates that the algorithms described can produce visually satisfactory reconstructions of the cardiac phantom from relatively small numbers of views, the most critical question is whether this can be achieved without hindering the detection of the small perfusion defects that is usually the goal of cardiac SPECT imaging. To answer this question we generated bullseye plots [26] from each set of reconstructions. To construct these plots, one starts with a set of short-axis slices of the left ventricle, which have the appearance of annuli. Each slice is mapped to a different ring of a dartboard-like grid divided into radial and azimuthal sectors, with the apex corresponding to the innermost ring (the bullseye) and the base corresponding to the outermost ring. The value of the integrated activity in a finite angular sector of each annulus is mapped to the appropriate azimuthal sector of the dartboard. If the reconstruction of the ventricle contained a uniform distribution of activity, then the bullseye plot would be uniform. But if the activity were non-uniform, for instance if there were a perfusion defect having lower activity than surrounding areas, then the appropriate sector of the dartboard would appear darker than the surrounding area. The bullseye plots corresponding to reconstructions from few-view sinograms that have been processed by the spline smoothing and interpolation techniques are shown in Fig. 3, along with bullseye plots corresponding to 120-angle sinograms reconstructed without spline processing. Our phantom contained a 1-cm perfusion defect insert, which produces a depression that is well resolved in many of the bullseye plots.

It is clear from these bullseye plots that the defect remains detectable for as few as 15 angles in the case of the 0-cm offset, as few as 30 angles in the case of the 9-cm offset, and as few as 60 angles in the case of the 15-cm offset. These findings correlate well with the visual appearances of the images in Fig. 2. The plots in which the defect is not visible correspond to the images in which severe interpolation artifacts are evident.

IV. DISCUSSION AND CONCLUSIONS

We have presented a sinogram pre-processing technique combining spline-based smoothing and interpolation that enables FBP to produce high-quality tomographic reconstructions from a smaller number of views than is usually used. The technique is applicable to situations where the number of views needed to satisfy the Nyquist condition on the sampling of the angular part of the sinogram is less than the number of views required by FBP to produce artifact-free images. In this situation, we first smooth each 1D projection
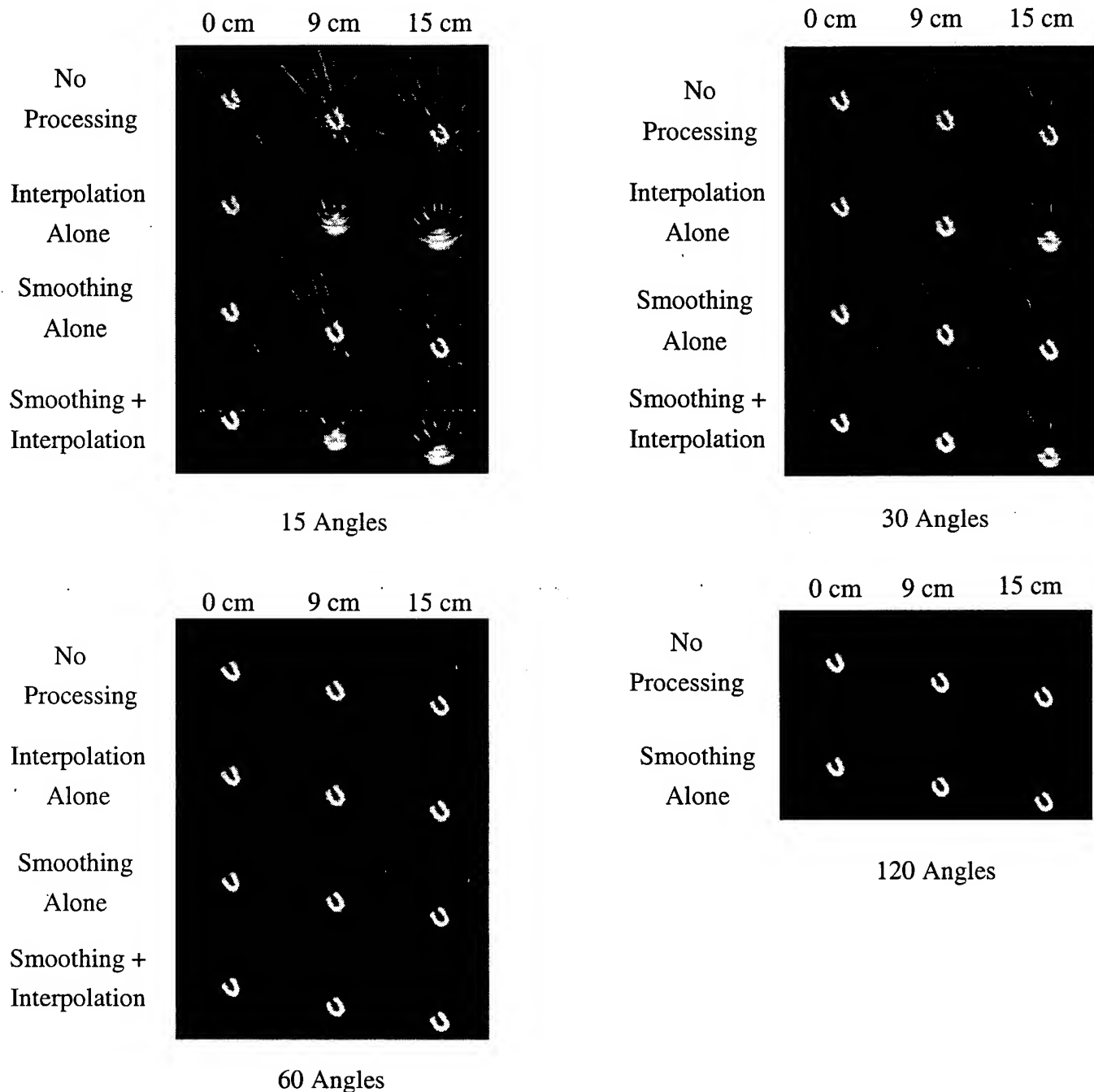
Figure 2: A representative slice of a cardiac phantom imaged at three different radial offsets and reconstructed with use of four different pre-processing approaches from 15, 30, 60, and 120 projection angles.

using a roughness-penalized nonparametric regression approach based on a GLM that explicitly models the Poisson statistics of the measured data. One-dimensional periodic interpolation splines are then fit in the angular direction and resampled to generate additional views. Ramp-filtered FBP is then applied to the interpolated sinogram.

Because it is not possible to derive closed-form expressions for the Nyquist and FBP sampling requirements of non-circularly symmetric objects, we have tested the limits of the algorithm by conducting an experimental investigation performing reconstructions from various numbers of projections of a cardiac phantom placed at various radial offsets from the center of rotation of a SPECT system. The ability of the algorithm to produce visually appealing images that still capture small perfusion defects breaks down for large radial offsets combined with small numbers of starting views. This corresponds to the situation when the number of initial angular samples strongly fails to satisfy the Nyquist condition for some or all projection bins and thus the assumptions underlying the approach are violated.
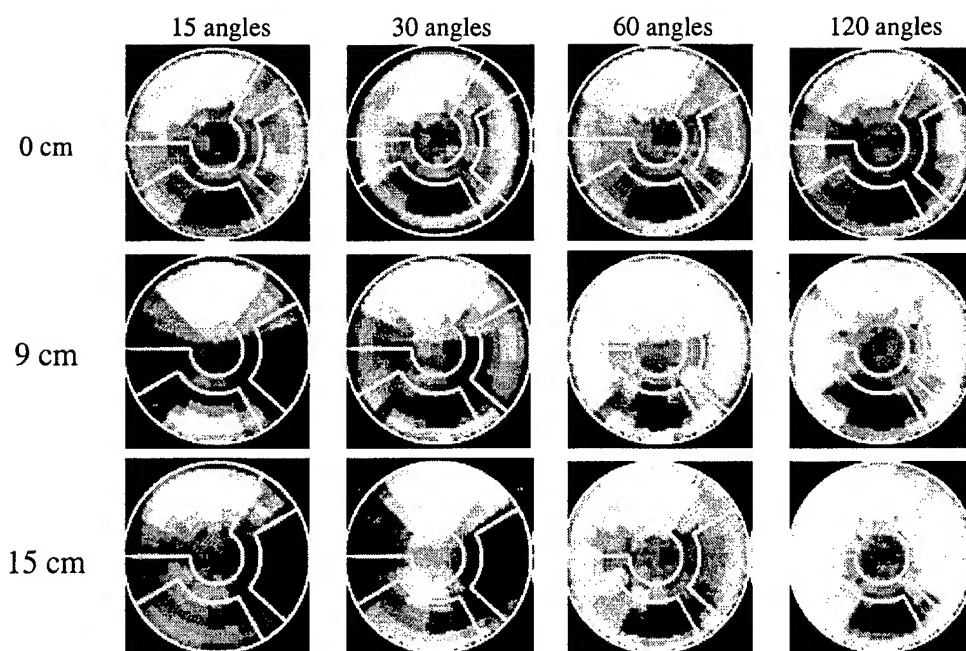
Figure 3: Bulleye plots for the cardiac phantom placed at various radial offsets and for four different numbers of projection angles. The reconstructions from 15, 30, and 60 angles used the nonparametric regression smoothing technique followed by periodic spline interpolation to 120 views prior to reconstruction by ramp-filtered FBP. The reconstructions from 120 angles simply entailed Hanning-filtered FBP.

V. ACKNOWLEDGMENT

VI. REFERENCES

[1] P. B. Heffernan and R. H. T. Bates, "Image reconstruction from projections. VI: Comparison of interpolation methods," *Optik*, vol. 60, pp. 129–142, 1982.

[2] R. A. Crowther, D. J. DeRosier, and A. Klug, "The reconstruction of a three-dimensional structure from projections and its application to electron microscopy," *Proc. Roy. Soc. Lond A*, vol. 317, pp. 319–340, 1970.

[3] K. A. Hanson and G. W. Wecksung, "Bayesian approach to limited-angle reconstruction in computed tomography," *J. Opt. Soc. Am.*, vol. 73, pp. 1501–1509, 1983.

[4] R. A. Brooks, G. H. Weiss, and A. J. Talbert, "A new approach to interpolation in computed tomography," *J. Comput. Assist. Tomogr.*, vol. 2, pp. 577–585, 1978.

[5] P. J. Green and B. W. Silverman, *Nonparametric Regression and Generalized Linear Models.* London: Chapman Hall, 1994.

[6] J. A. Fessler, "Tomographic reconstruction using information-weighted spline smoothing," in *Information Processing in Medical Imaging* (H. H. Barrett and A. F. Gmitro, eds.), pp. 290–300, Berlin: Springer-Verlag, 1993.

[7] S. X. Pan and A. C. Kak, "A computational study of reconstruction algorithms for diffraction tomography: Interpolation versus filtered backpropagation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 31, pp. 1262–1275, 1983.

[8] H. Stark, J. W. Woods, I. Paul, and R. Hingorani, "Direct Fourier reconstruction in computer tomography," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 29, pp. 237–245, 1981.

[9] H. Stark, J. W. Woods, I. Paul, and R. Hingorani, "An investigation of computerized tomography by direct Fourier inversion and optimum interpolation," *IEEE Trans. Biomed. Eng.*, vol. 28, pp. 496–505, 1981.

[10] P. R. Smith, T. M. Peters, and R. H. T. Bates, "Image reconstruction from a finite number of projections," *J. Phys. A*, vol. 6, pp. 361–382, 1973.

[11] H. Stark, "Sampling theorems in polar coordinates," *J. Opt. Soc. Am.*, vol. 69, pp. 1519–1525, 1979.

[12] H. Stark and M. Wengrovitz, "Comments and corrections on the use of polar sampling theorems in CT," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 31, pp. 1329–1331, 1983.

[13] D. Fraser, "Interpolation by the FFT revisited—an experimental investigation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 665–675, 1989.

[14] T. J. Cavicchi, "DFT time-domain interpolation," *IEE Proceedings-F*, vol. 139, pp. 207–211, 1992.

[15] P. J. La Rivière and X. Pan, "Mathematical equivalence of zero-padding and circular sampling thorem interpolation with implications for direct Fourier image reconstruction," in *Proc. SPIE*, vol. 3338, pp. 1117–1126, 1998.

[16] E. O. Brigham, *The Fast Fourier Transform and Its Applications*. Englewood Cliffs: Prentice Hall, 1988.

[17] P. J. La Rivière and X. Pan, "Comparison of angular interpolation approaches in few-view tomography using statistical hypothesis testing," in *Proc. SPIE*, vol. 3661, 1999. (in press).

[18] L. A. Shepp and B. F. Logan, "The Fourier reconstruction of a head section," *IEEE Trans. Nucl. Sci.*, vol. 21, pp. 21–43, 1974.

[19] H. Spath, *One-Dimensional Spline Interpolation Algorithms*. Wellesley, MA: A. K. Peters, 1993.

[20] J. A. Rice, *Mathematical Statistics and Data Analysis*. Belmont, CA: Wadsworth, 1994.

[21] P. McCullagh and J. A. Nelder, *Generalized Linear Models*. London: Chapman Hall, 1983.

[22] N. P. Galatsanos and A. K. Katsaggelos, "Method for choosing the regularization parameter and estimating the noise variance in image restoration and their relation," *IEEE Tr. Im. Proc.*, vol. 1, pp. 322–336, 1992.

[23] G. Wahba, *Spline Models for Observational Data*. Philadelphia: SIAM Press, 1990.

[24] F. O'Sullivan, B. S. Yandell, and W. J. Raynor, Jr., "Automatic smoothing of regression functions in generalized linear models," *J. Amer. Stat. Assoc.*, vol. 81, pp. 96–103, 1986.

[25] M. F. Hutchinson and F. R. de Hoog, "Smoothing noisy data with spline functions," *Numer. Math.*, vol. 47, pp. 99–106, 1985.

[26] E. V. Garcia, K. V. Train, J. Maddahi, F. Prigent, J. Friedman, J. Areeda, A. Waxman, and D. S. Berman, "Quantification of rotational thallium-201 myocardial tomography," *J. Nucl. Med.*, vol. 26, pp. 17–26, 1985.

A Fourier-Based Optimal Recovery Method for Anti-Aliasing Interpolation

Chien-Min Kao, Xiaochuan Pan, Patrick La Rivière, and Mark A. Anastasio

Department of Radiology
University of Chicago
5841 S. Maryland Ave., MC-1037
Chicago, Illinois 60637
Phone: 773-702-6273
FAX: 773-702-5986
Email: c-kao@uchicago.edu

Abstract

We propose a Fourier-based interpolation method designed to suppress the oscillatory artifacts that occur when interpolating an undersampled function. In the proposed method, we formulate the problem as an optimal recovery problem incorporating *a priori* knowledge of the aliasing. Our computer simulations demonstrate that the proposed method can produce good interpolation results even when the functions are severely undersampled.

Subject terms: anti-aliasing interpolation; optimal recovery; FFT; image enlargement.

1 Introduction

Interpolation is an important operation in digital signal and image processing. For example, it is used to change the sampling rate in speech processing [1] and to enlarge images [2]. In signal processing, the Whittaker-Shannon interpolation theorem has established the use of ideal lowpass filtering for the interpolation of a bandlimited function from an infinite, equispaced sequence of its samples [3]. In the spatial domain, this leads to the use of a sinc kernel for ideal interpolation [3]. However, this interpolation scheme is seldom used because of the great computation burden involved in calculating the slowly decaying sinc function. Instead, most practical interpolation methods, including nearest-neighbor interpolation, linear interpolation and cubic B-spline interpolation, use short interpolation kernels for fast implementation. For improved interpolation, finite-impulse-response (FIR) filters that approximate the ideal lowpass interpolation filter can be designed [4]. An alternative approach is to fit the samples with functions of predetermined forms, such as polynomial functions [5] and splines [6].

The development of the fast Fourier transform (FFT) algorithm [7] has given rise to algorithms based on the use of discrete Fourier transform (DFT) [8, 9, 10]. In this method, the DFT of the sequence under study is padded with zeros to a desired length and a more densely sampled, interpolated sequence obtained by taking the inverse DFT of the zero-padded sequence. It has been shown recently that this approach is exact for bandlimited, periodic functions [10], and that it is also a computationally efficient implementation of the circular sampling theorem (CST), a counterpart of the Whittaker-Shannon interpolation for periodic functions [11].

However, in many situations of practical interest, the functions are not bandlimited exactly to the Nyquist frequency of the sampling. Consequently, aliasing occurs due to undersampling of the functions. When interpolated by DFT zero-padding, the aliasing errors appear in the form of oscillatory artifacts. It is the intent of this work to develop a DFT-based method for interpolation of undersampled functions that can mitigate such oscillatory artifacts. This will be achieved by incorporating *a priori* knowledge of the aliasing in an optimal recovery approach formulated in the discrete frequency space.

This paper is organized as follows. In Sec. II, a DFT-based method for interpolation will be developed from the perspective of optimal recovery. In Sec. III, we will conduct several computer experiments to demonstrate the performance of the proposed method. Finally, conclusion and discussions will be presented in Sec. IV.

2 Theory

In this paper, we will denote by $\{a_n\}_N$ a finite sequence $\{a_0, a_1, \ldots, a_{N-1}\}$ of N elements. Because the method developed below is based on DFT which implicitly assumes the functions it works with are periodic, the finite sequence $\{a_n\}_N$ will be assumed to be part of a periodic sequence $\{a_n\}_\infty$ satisfying $a_{n+kN} = a_n$, $n = 0, \ldots, N-1$, $k \in Z$.

Given a sequence of samples $\{f_n\}_N$, we consider the interpolation problem of finding a new sequence $\{g_m\}_M$, $M = PN$, $P > 1$, such that

$$g_{nP} = f_n, \ n = 0, \ldots, N-1. \tag{1}$$

Let $\{G_m\}_M$ and $\{F_n\}_N$ be the DFT sequences of $\{g_m\}_M$ and $\{f_n\}_N$, respectively. By definition, we have $f_n = \sum_{n'=0}^{N-1} F_{n'} \, e^{j2\pi n n'/N}$. On the other hand, $g_{nP} = \sum_{m=0}^{PN-1} G_m \, e^{j2\pi m n P/M} = \sum_{n'=0}^{N-1} (\sum_{p=0}^{P-1} G_{n'+pN}) \, e^{j2\pi n n'/N}$. Therefore, Eq. (1) is equivalent to

$$\sum_{p=0}^{P-1} G_{n+pN} = F_n, \quad n = 0, \dots, N-1 \tag{2}$$

in the discrete frequency space. It is worth noting that Eq. (2) is the discrete counterpart of the Whittaker-Shannon sampling theorem, which states that $F(\nu) = \sum_{n=-\infty}^{\infty} G(\nu - n\nu_N)$, where $F(\nu)$ and $G(\nu)$ are the spectra of the sampled and the original functions, respectively, and ν_N is the Nyquist frequency of the sampling.

The N constraints given by Eq. (2) are insufficient to determine the M values of G_m uniquely. To obtain a unique solution, we define a cost function of the form

$$\Phi_{\{G_m\}} = \sum_{m=0}^{M-1} \left\| G_m - \bar{G}_m \right\|^2 + \beta \sum_{m=0}^{M-1} \omega_m \left\| G_m \right\|^2, \tag{3}$$

where $\{\bar{G}_m\}_M$ is the DFT of the *a priori* solution $\{\bar{g}_m\}_M$, β is a positive real number, and $\{\omega_m\}_M$ defines a discrete filter such that the second summation is a global roughness measure of $\{g_m\}_M$. The solution $\{G_m\}_M$ is then obtained by minimizing $\Phi_{\{G_m\}}$ subject to the condition of Eq. (2). Thus, in spirit of signal recovery [12], the interpolation is achieved by finding the sequence that passes through the measured points $\{f_n\}_N$ having the best tradeoff between resemblance to the *a priori* solution $\{\bar{g}_m\}_M$ and a roughness penalty. The tradeoff is controlled by the hyperparameter β: a larger β will force to produce smoother solution while a smaller β will yield a solution closer to the *a priori* solution.

As shown in the Appendix, the solution to the minimization problem is given by

$$G_{n+pN} = \alpha_{n+pN} \left\{ F_n + \gamma_n \Delta \bar{G}_{n+pN} \right\}, \tag{4}$$

where $\alpha_{n+pN} = \gamma_n^{-1}(1+\beta\omega_{n+pN})^{-1}$, $\gamma_n = \sum_{p=0}^{P-1}(1+\beta\omega_{n+pN})^{-1}$, and $\Delta\bar{G}_{n+pN} = \bar{G}_{n+pN} - \sum_{p=0}^{P-1} \alpha_{n+pN}\bar{G}_{n+pN}$. It should be emphasized again that because Eq. (4) satisfies Eq. (2), regardless of the *a priori* solution used, the solution obtained is always an interpolating one, i.e., it always contains $\{f_n\}_N$ as a subsequence. This property sets it apart from other anti-aliasing interpolation methods involving lowpass smoothing, in which the results do not pass through all the sample points.

A. Choice of the Roughness Filter $\{\omega_m\}_M$

In choosing $\{\omega_m\}_M$, it is desirable for Eq. (4) to yield the zero-padding interpolation (i.e., the CST interpolation) given by

$$G_m = \begin{cases} F_m & 0 \leq m \leq \lfloor \frac{N}{2} \rfloor \\ 0 & \lfloor \frac{N}{2} \rfloor + 1 \leq m \leq M - \lfloor \frac{N}{2} \rfloor \\ F_{m-M+N} & M - \lfloor \frac{N}{2} \rfloor \leq m \leq M - 1, \end{cases} \tag{5}$$

when $\beta \to \infty$ and $\Phi_{\{G_m\}}$ depends on $\sum_{m=0}^{M-1} \omega_m \left\| G_m \right\|^2$. It is straightforward to see that this condition results in a $\{\omega_m\}_M$ which is zero except for $m = \lfloor \frac{N}{2} \rfloor + 1, \dots, M - \lfloor \frac{N}{2} \rfloor$. On the other hand, a widely used roughness measure of a sequence $\{g_m\}_M$ is the energy of its second-order derivative [12, 13], which in the discrete frequency space is proportional to $\sum_{m=0}^{M-1} \sin^4\left(\frac{\pi m}{M}\right) \left\| G_m \right\|^2$. By taking both considerations into account, we thus define

$$\omega_m = \begin{cases} \sin^4(\frac{\pi m}{M}) & \lfloor \frac{N}{2} \rfloor + 1 \leq m \leq M - \lfloor \frac{N}{2} \rfloor \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

B. Construction of $\{\bar{g}_m\}_M$

Generally, the construction of $\{\bar{g}_m\}_M$ should be based on *a priori* knowledge of the function under study, thus allowing partial recovery of the information lost in the sampling. In this work, however, we will consider a more limited scope of applications for Eq. (4), i.e., we will consider its use for suppression of the oscillatory aliasing artifacts due to undersampling of the function, rather than recovery of the lost information, when no *a priori* knowledge is available.

For this application, the design of $\{\bar{g}_m\}_M$ is based on the following observations. First, with respect to smoothness and aliasing errors, linear interpolation usually generates adequate results except for the slope discontinuities at the sample points. On the other hand, using DFT zero-padding interpolation, continuous slopes are obtained, but aliasing errors are often observed. These aliasing errors typically occur in the form of overshooting at sharp transitions, resulting in oscillatory artifacts in flat regions. Consider the simple example shown in Fig. 1. For the sample point f_n, there are two line segments, l_{n-1} and l_n, connecting to it from its two nearest neighbors f_{n-1} and f_{n+1}, respectively. When overshooting occurs at f_n, it typically happens to the line segment having smaller absolute slope, which is l_{n-1} in our example. Thus, one way to suppress the overshooting is to purposely place some points of the *a priori* function in the interval of l_{n-1} to the opposite side of the line segment as the overshoot, such as the point P_1 in Fig. 1.

Based on these observations, we construct the *a priori* function as follows. Except for the two nearest neighbors of the sampled points, $\{\bar{g}_m\}_M$ is obtained by linear interpolation, i.e.,

$$\bar{g}_{nP+j} = g_{nP+j}^{(linear)} = f_n + \frac{j}{P}(f_{n+1} - f_n),$$
(7)

for $j = 0, \ldots, P-1, j \neq 1, P-1$, and $n = 0, \ldots, N-1$.

Next, at each sample point, we compare the absolute slopes s_{n-1} and s_n of the line segments l_{n-1} and l_n, respectively. If $s_{n-1} < s_n$, overshooting occurs on l_{n-1}. Hence, we define

$$\bar{g}_{nP-1} = g_{nP-1}^{(linear)} + dg_n,$$
(8)

where dg_n is a number large enough to balance the trend of overshooting at f_n. Consider the extreme case of overshooting in which the line segment l_n is continued to the point f_{n-1}. As an estimate, we can use the difference between the value f_{n-1} and the value of the extension of l_n to this point, $f_n - (f_{n+1} - f_n) = 2f_n - f_{n+1}$, as a measure of the "strength" of the overshooting. Hence, we define $dg_n = f_{n-1} - (2f_n - f_{n+1}) = f_{n+1} - 2f_n + f_{n-1}$, which is the discrete second-order derivative of f_n. To help smooth transition at f_n, we also set

$$\bar{g}_{nP+1} = g_{nP+1}^{(linear)} - dg_n.$$
(9)

In the case $s_{n-1} > s_n$, overshooting occurs on l_n. By applying the same consideration outlined above, we obtain

$$\bar{g}_{nP\pm1} = g_{np\pm1}^{(linear)} \pm dg_n.$$
(10)

3 Results

To examine the performance of the proposed method, three data sets were simulated and interpolated. Each data set was obtained by sampling an analytic function defined on the interval $[0, 1]$ with a sampling distance of $1/16$, thus generating 16 samples for each set. These samples were then interpolated by use of various methods to improve the sampling rate by a factor of 16, thereby generating 256 samples for each set after interpolation.

Figure 2 compares the interpolation results of a rectangular function and of an impulse-like function, $f(x) = 1/[1 + (\frac{x-0.5}{0.02})^2]$, by using the proposed method with $\beta = 1.2 \times 10^2$ (Proposed), the zero-padding method (ZeroPad), and the cubic B-spline interpolation (Cubic-B). The oscillatory aliasing errors can be clearly observed in both zero-padding and cubic B-spline results. These artifacts are effectively suppressed using the proposed method.

Figure 3 shows the results obtained from samples of a function consisted of sums of Gaussian functions. In this case, the results obtained by all three methods are significantly different from the actual function. This is not surprising because the function is severely undersampled. However, the result generated by the proposed method (using $\beta = 1.2 \times 10^2$) appears to conform quite well to the limited information provided by the samples without generating noticeable oscillatory artifacts.

We also applied the proposed one-dimensional (1D) interpolation method for enlargement of a two-dimensional (2D) $N \times N$ image to a $M \times M$ image where $M = PN, P > 1$. This is achieved by a row-by-row interpolation of the $N \times N$ image to produce a $N \times M$ image, followed by a column-by-column interpolation of the $N \times M$ image to generate the final $M \times M$ image. Figure 4 shows the enlarged 256×256 images from a 64×64 MRI brain image using various methods. Aliasing artifacts can clearly be observed in the image obtained by DFT zero-padding (Fig. 4(a)). These artifacts are significantly reduced in the image obtained by the proposed method (Fig. 4(b)). Nearest-neighbor interpolation is computationally very simple, but the result is not visually appealing (Fig. 4(c)). In comparison with the image obtained by bilinear interpolation (Fig. 4(d)), the image generated by the proposed method appears to have better resolution and contrast.

4 Conclusions and Discussion

In this paper, a DFT-based method for 1D anti-aliasing interpolation was developed from the perspective of optimal recovery. The proposed method is not based on function approximation. Therefore, it is particularly suited for interpolation problems in which no accurate mathematical modeling is available for the underlying functions generating the samples. Many interpolation problem of practical importance, such as image enlargement, belongs to this class. In this work, 2D image enlargement was achieved by interpolating each dimension in turn. Good results were obtained in this way. Although this method is not optimal because it does not make use of the full 2D information available in the data, it can readily be extended to use the 2D FFT for 2D interpolation.

Appendix

To minimize $\Phi_{\{G_m\}}$ given by Eq. (3) subject to Eq. (2), we define

$$\Psi_{\{G_m\}} = \Phi_{\{G_m\}} + \sum_{n=1}^{N-1} \lambda_n \left\| F_n - \sum_{p=0}^{P-1} G_{n+pN} \right\|^2, \tag{11}$$

where λ_n's are some positive numbers. Taking derivative of $\Psi_{\{G_m\}}$ with respect to G^*_{n+pN} (the complex conjugate of G_{n+pN}) and equating it to zero, we obtain

$$\varpi_{n+pN} G_{n+pN} + \lambda_n \sum_{p=0}^{P-1} G_{n+pN} = \lambda_n F_n + \bar{G}_{n+pN}, \tag{12}$$

where $\varpi_m = 1 + \beta\omega_m$, $m = 0, \dots, M-1$, for $n = 0, \dots, N-1$, $p = 0, \dots, P-1$. It follows from Eq. (12) that

$$(1 + \lambda_n\gamma_n) \sum_{p=0}^{P-1} G_{n+pN} = \lambda_n\gamma_n F_n + \sum_{p=0}^{P-1} \varpi_{n+pN}^{-1} \bar{G}_{n+pN}, \tag{13}$$

where $\gamma_n = \sum_{p=0}^{P-1} \varpi_{n+pN}^{-1}$. Substitution of Eq. (13) into Eq. (12) yields

$$\varpi_{n+pN} G_{n+pN} = \frac{\lambda_n}{1 + \lambda_n\gamma_n} F_n + \bar{G}_{n+pN} - \frac{\lambda_n}{1 + \lambda_n\gamma_n} \sum_{p=0}^{P-1} \varpi_{n+pN}^{-1} \bar{G}_{n+pN}. \tag{14}$$

The solution that minimizes $\Phi_{\{G_m\}}$, Eq. (4), is then given by the limiting solution of Eq. (14) when $\lambda_n \to \infty, \forall n$.

References

[1] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, Bell Laboratories, Inc., New Jersey (1978).

[2] M. Unser, A. Aldroubi and M. Eden, "Enlargement or reduction of digital images with minimum loss of information," *IEEE Trans. Image Processing* **4**, 247-258 (1995).

[3] R. E. Crochiere and L. R. Rabiner, "Interpolation and decimation of digital signals - A Tutorial Review," *Proc. IEEE* **69**, 300-334 (1981).

[4] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, Prentice Hall, Englewood Cliffs, New Jersey (1989).

[5] P. J. Davis, *Interpolation and Approximation*, Blaisdell Publishing Company, Waltham, Massachusetts (1963).

[6] H. Späth, *One Dimensional Spline Interpolation Algorithms*, A K Peters Ltd., Massachusetts (1995).

[7] E. O. Brigham, *The Fast Fourier Transform and Its Applications*, Prentice Hall, New Jersey (1988).

[8] K. P. Prasad and P. Satyanarayana, "Fast interpolation algorithm using FFT," *Electron. Lett.* **22**, 185-187 (1986).
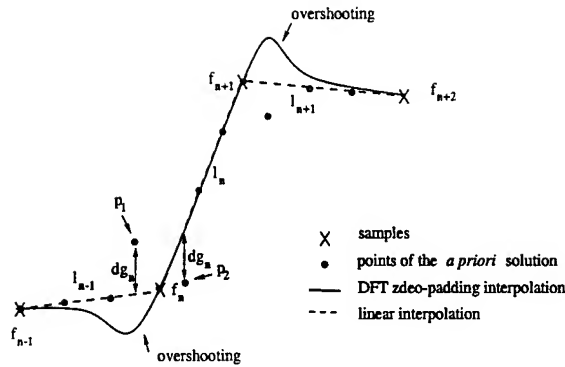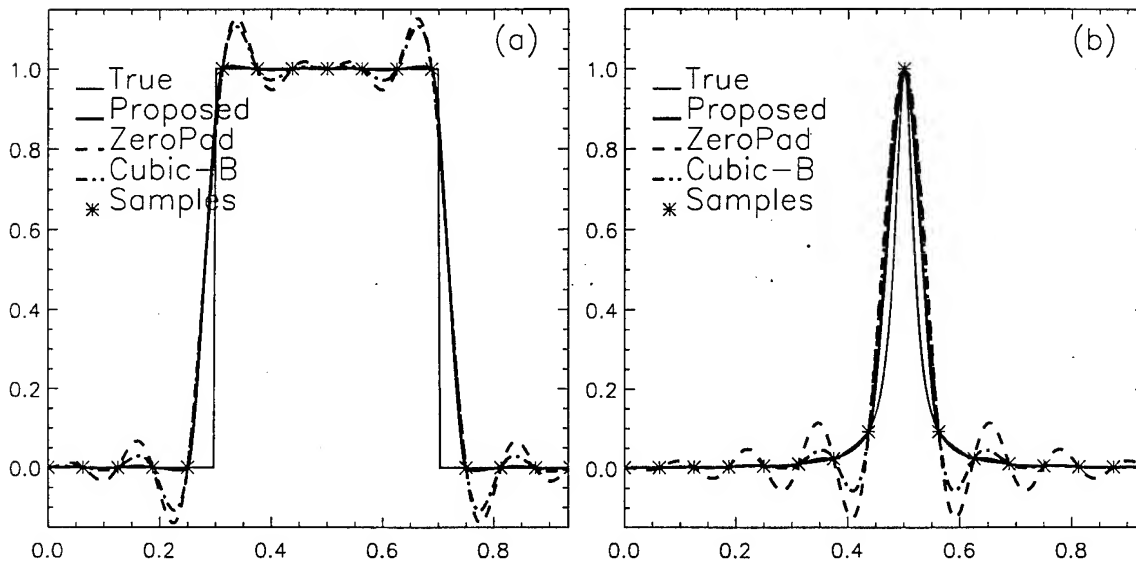
Figure 1: Overshooting in interpolation.

Figure 2: Interpolation of samples of (a) a rectangular function and (b) an impulse-like function defined by $f(x) = 1/[1 + (\frac{x-0.5}{0.02})^2]$ using various methods.

[9] T. J. Cavicchi, "DFT time-domain interpolation," *IEE Proceedings-F* **139**, 207-211 (1992).

[10] P. La Rivière and X. Pan, "Mathematical equivalence of zero-padding interpolation and circular sampling theorem interpolation with implications for direct Fourier image reconstruction," *Proc. SPIE* **3338**, K.M. Hanson Ed., pp. 1117-1126, SPIE Press, Bellingham, WA (1998).

[11] H. Stark, "Sampling theorems in polar coordinates," *J. Opt. Soc. Am.* **69**, 1519-1525 (1979).

[12] G. Dermont, "Image reconstruction and restoration: overview of common estimation structures and problems," *IEEE Trans. Acoust., Speech, Signal Processing* **37**, 2024-2036 (1989).

[13] D.M. Titterington, "General structure of regularization procedures in image reconstruction," *Astron. Astrophys.* **144**, 381-387 (1985).
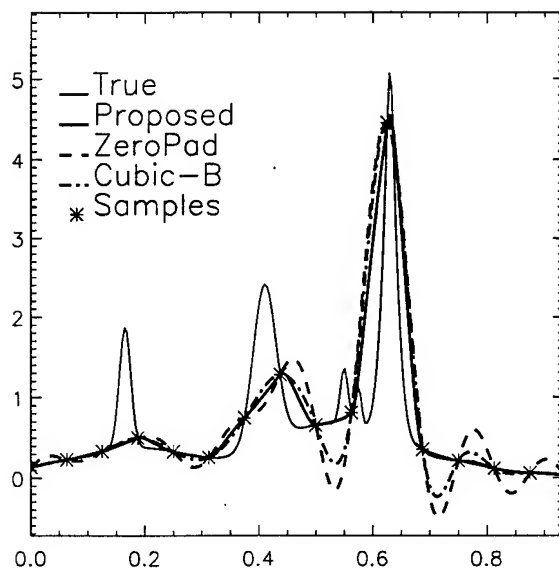
Figure 3: Interpolation of samples of a function consisted of sums of Gaussian functions.
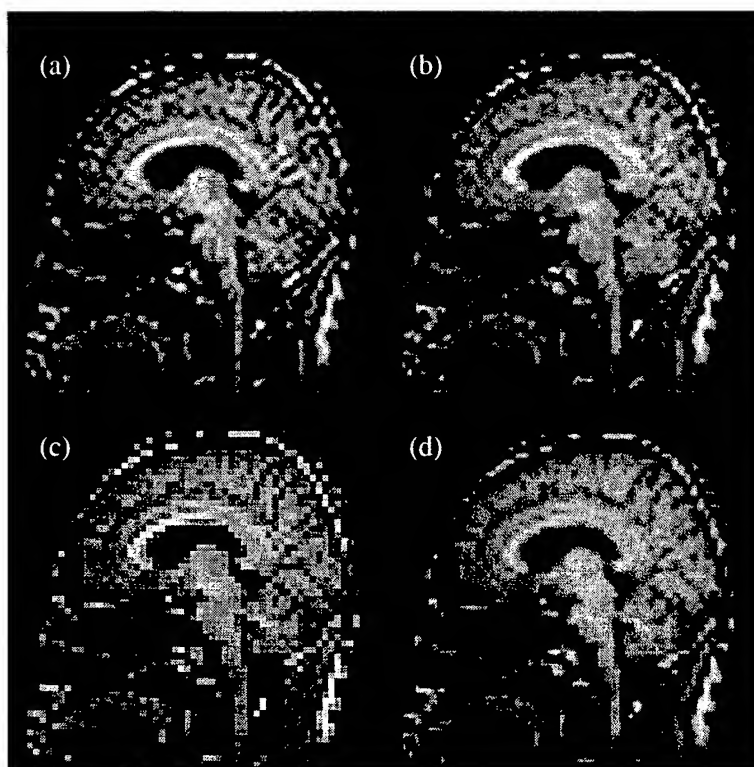

Figure 4: Image enlargement using (a) DFT zero-padding, (b) the proposed method, (c) nearest-neighbor interpolation, and (d) bilinear interpolation.

Ideal-Observer Analysis of Lesion Detectability in Planar, Conventional SPECT, and Dedicated SPECT Scintimammography Using Effective Multi-Dimensional Smoothing

P.J. La Rivière, *Student Member, IEEE*, X. Pan, *Member, IEEE*, and B.C. Penney

Department of Radiology, The University of Chicago, Chicago, IL, 60637

Abstract

Scintimammography, a nuclear-medicine imaging technique that relies on the preferential uptake of Tc-99m-sestamibi and other radionuclides in breast malignancies, has the potential to provide differentiation of mammographically suspicious lesions, as well as outright detection of malignancies in women with radiographically dense breasts. In this work we use the ideal-observer framework to quantify the detectability of a 1-cm lesion using three different imaging geometries: the planar technique that is the current clinical standard, conventional single-photon emission computed tomography (SPECT), in which the scintillation cameras rotate around the entire torso, and dedicated breast SPECT, in which the cameras rotate around the breast alone. We also introduce an adaptive smoothing technique for the processing of planar images and of sinograms that exploits Fourier transforms to achieve effective multidimensional smoothing at a reasonable computational cost.

For the detection of a 1-cm lesion with a clinically typical 6:1 tumor-background ratio, we find ideal-observer signal-to-noise ratios (SNR) that suggest that the dedicated breast SPECT geometry is the most effective of the three, and that the adaptive, two-dimensional smoothing technique should enhance lesion detectability in the tomographic reconstructions.

I. INTRODUCTION

Breast cancer is the most frequently diagnosed invasive malignancy among American women and ranks second only to lung cancer in annual cancer-related mortality for this group [1]. Numerous studies have shown that early detection and treatment of breast cancer can improve survival rates [2-4]. Screen-film mammography has come to play a vital role in this detection process, due to its high (80-90%) sensitivity to breast malignancies. However, mammography is notoriously poor at distinguishing benign from malignant tumors, having reported specificities and positive predictive values of 15-30% [5]. This means that only a small percentage of lesions biopsied on the basis of suspicious mammographic appearance are found to be malignant.

In recent years, researchers have developed and studied a nuclear-medicine test with the potential to provide relatively low-cost, minimally invasive differentiation of breast abnormalities identified by physical examination or mammography [6-18]. Known as scintimammography, the test relies on the preferential uptake of Tc-99m-sestamibi or other radionuclides such as Tl-201, Tc-99m- tetrofosmin, or Tc-99m-MDP in breast malignancies as compared to normal breast tissue or benign abnormalities. Indeed, one study has shown that typical *in vivo* tumor-background concentration ratios of Tc-99m-sestamibi are 5.64±3.06 [19]. This focal uptake can be imaged in a number of ways, though the most widely used clinical protocol involves acquiring one or two planar views—one lateral view and possibly an additional oblique or anterior view—while the patient lies prone on a specially designed table [20]. The imaging time is typically 10-15 minutes per view. Numerous clinical studies with histological follow-up have been performed using this or a similar protocol, reporting sensitivities of 83-96%, and specificities of 66-100% when using Tc-99m-sestamibi to image mammographically suspicious lesions [8-18]. In addition to differentiating breast abnormalities detected by other means, scintimammography may also have a role in the detection of breast malignancies in patients with radiographically dense breasts, for whom screen-film mammograms are often difficult to interpret [7].

A few of these studies of scintimammography have also examined the role of conventional SPECT (where the patient lies supine and the camera circles the torso) in detecting focal uptake of Tc-99m-sestamibi and have found comparable but not generally improved sensitivities as compared to planar techniques, along with substantially lower specificities [14,21-23]. Wang *et al.* [24] speculated that this surprisingly poor performance was due to substantial attenuation of photons emitted in the breast by the torso in at least half of the views as well as to the presence of scatter from organs such as the heart and liver known to have high uptake of Tc-99m-sestamibi. The poor performance of conventional SPECT may also be related to the inferior resolution of conventional SPECT as compared to planar techniques in this situation, due to the fact that the scintillation cameras are on average further away from the breast in the conventional SPECT geometry than in the planar geometry. Wang *et al.* also investigated a geometry that they called vertical axis-of-rotation SPECT and that we call dedicated breast SPECT, or simply dedicated SPECT, in which the scintillation cameras are assumed to rotate around one breast alone. This geometry eliminates the effect of attenuation by the thorax and, with proper shielding, the effect of scatter from the thorax. Moreover, the small radius of rotation offers improved resolution and sensitivity. In phantom studies, Wang *et al.* found that with this dedicated geometry they were able to detect a breast lesion with an outer diameter of 1 cm and a 6:1 lesion-to-background concentration ratio that was not detectable in either conventional SPECT or planar studies with the same total imaging time.

The present work examines quantitatively the question of lesion detectability in these three different geometries—planar, conventional SPECT, and dedicated SPECT—using the so-called ideal-observer framework to calculate signal-to-noise ratios as a function of lesion concentration for the different geometries, and, in the case of the tomographic geometries, for different reconstruction filters. It also

introduces and applies an adaptive smoothing technique for the processing of planar images or sinograms that exploits Fourier transforms to achieve effective multidimensional smoothing at a reasonable computational cost. Such pre-processing is found to improve idealized lesion detectability in the reconstructed images.

II. METHODS

A. Effective multi-dimensional smoothing

The projection data acquired by tomographic nuclear-medicine imaging devices is invariably contaminated by noise, which is propagated and often amplified by the reconstruction algorithm. Filtered backprojection (FBP), the most computationally efficient and most commonly used of these algorithms, attempts to control this noise by incorporating an apodization window, such as a Hanning filter, into the projection filtration step. Alternatively, in some reconstructions, only the mathematically exact ramp filter is applied to the projection data, but the reconstructed image is subjected to a two- (or three-) dimensional post-filtering. While somewhat different in application and effect, these two approaches have the common property of being linear; the degree of smoothing applied to the data does not depend on the data, but is rather fixed a priori, and this same degree of smoothing is applied uniformly to the entire dataset.

In contrast, non-linear, adaptive smoothing methods, such as the generalized cross-validation method to be discussed below, determine the degree of smoothing to be applied to various specified subsets of the data from the statistics of each such subset and the degree of smoothing can vary from subset to subset. For instance, in 2D image reconstruction, the data at each projection angle could be smoothed differently, with the degree of smoothing determined from the statistics of the data at each angle.

When smoothing for 2D image reconstruction, one would in principle like to exploit the statistical correlations between different projections as well as those within a given projection. A truly 2D adaptive smoothing operation of this type can be computationally expensive and difficult to implement [25]. However, by exploiting the properties of the Fourier transform, one can achieve effective 2D smoothing at the cost of a series of 1D smoothing operations. Consider a 2D discrete sinogram $p(\xi_i, \theta_j)$, where ξ_i refers to the ith projection bin $(i=1,...,N)$ and θ_j the jth projection angle $(j=1,...M)$. It can be shown [26] that the following sequence of operations is equivalent to an adaptive, 2D smoothing of the sinogram:

1. Take a 1D discrete Fourier transform of the sinogram with respect to the projection angle θ_j; the result can be viewed as a set of 1D functions of the untransformed variable ξ_i, each labeled by an angular frequency index k, i.e. $P_k(\xi_i)$.

2. Perform an adaptive 1D smoothing of each of these M functions of ξ_i, yielding M discrete smoothed functions $P_k^s(\xi_i)$, where the superscript s stands for smoothed.

3. Perform an inverse 1D discrete Fourier transform of $P_k^s(\xi_i)$ with respect to the angular frequency k to recover $p^s(\xi_i, \theta_j)$.

The adaptive 1D smoothing we perform on each of the M functions $P_k(\xi_i)$ is known as penalized least-squares smoothing [27,28], and involves fitting the discrete data with a continuous smoothing curve $P_k^s(\xi)$ that minimizes the functional

$$S\{P_k^s(\xi)\} = \sum_{i=1}^N \left[P_k(\xi_i) - P_k^s(\xi_i) \right]^2 + \alpha \int_0^T [(P_k^s(\xi))'']^2 d\xi, \quad (1)$$

where ξ is a continuous variable representing the position along a given projection, T is the total length of the projection, and the double prime denotes the second derivative with respect to ξ. The two terms in this functional represent the competing goals of achieving a good fit to the data while maintaining a smooth curve, with the parameter α mediating the tradeoff. For instance, if α is zero, the smoothness constraint disappears and the minimizing curve will be a piecewise linear interpolant to the data; if α grows large, the smoothness constraint dominates and the curve approaches a simple linear fit to the data. For intermediate values of α, the minimizing curves balance the goodness-of-fit and smoothness constraints. It can in fact be shown that the minimizers of this functional will always be members of a class of functions known as natural cubic splines [27,28]. These are piecewise cubic curves that join at the abscissa values ξ_i, where they are continuous up to and including second derivative.

Clearly the choice of α determines the degree of smoothing that is applied to the data, and it is this parameter that is determined from the statistics of the data itself in an adaptive implementation of penalized least-squares smoothing using an algorithm known as generalized cross-validation [29]. Thus a generally different α is used in the smoothing of each of the M functions $P_k(\xi_i)$. The resulting continuous smoothed functions $P_k^s(\xi)$ must then be sampled to yield the discrete functions $P_k^s(\xi_i)$ which are subjected to the inverse DFT in step 3 above.

It should also be noted that while we have, for the sake of simplicity, discussed effective 2D smoothing, the technique can be extended to any number of dimensions. To smooth an n-dimensional function, one can simply take an $(n-1)$-dimensional Fourier transform of the function and then perform a set of 1D smoothings over the untransformed variable prior to taking an inverse $(n-1)$-dimensional Fourier transform [26].

B. Ideal Observer Framework

The ideal-observer framework [30] offers a way of assessing the amount of information the data from an imaging device contain with regard to the performance of a specified task. For example, the simplest such task is the detection of a signal of known strength, shape, and location in a specified background. In this case, the framework seeks to quantify the degree to which an ideal observer—one who can use the information contained in the images to its fullest extent—can reliably distinguish images containing the

background alone from images containing the background and the signal when both kinds of images are corrupted by noise, blurring, and other imperfections. For linear imaging processes in which the noise in the output image is assumed to be additive, Gaussian, zero-mean, stationary, and independent of the presence or absence of the signal, the ideal-observer framework allows us to characterize fully the quality of the imaging system data with respect to the performance of the specified signal-detection task by a single number, the ideal-observer signal-to-noise ratio (SNR). This is usually expressed as

$$SNR_I^2 = K^2 \int \frac{|\Delta S_{in}(\upsilon)|^2 MTF^2(\upsilon)}{W(\upsilon)} d\upsilon, \qquad (2)$$

where K is the large-scale transfer characteristic of the imaging system at the desired operating point, $MTF(\upsilon)$ is the system modulation transfer function, $W(\upsilon)$ is the system Wiener spectrum and $|\Delta S_{in}(\upsilon)|^2$ is the power spectrum of the signal in input space, i.e. before it has been scaled and degraded by the imaging system. This expression allows one to determine the ideal-observer SNR for *any* analytically specified input signal once K, $MTF(\upsilon)$, and $W(\upsilon)$ are known. Alternatively, if one wishes to determine the ideal-observer SNR for a particular real signal, equation (2) could be re-expressed as

$$SNR_I^2 = \int \frac{|\Delta S_{out}(\upsilon)|^2}{W(\upsilon)} d\upsilon, \qquad (3)$$

where $|\Delta S_{out}(\upsilon)|^2$ is the power spectrum of the signal in output space, i.e. after it has been scaled and degraded by the imaging system. This is the form we will use, because by acquiring an ensemble of images containing the signal and the background as well as an ensemble of images containing the background alone, $|\Delta S_{out}(\upsilon)|^2$ can be easily determined by computing the power spectrum of the difference between the two ensemble averages.

Finally, when the noise in an image is uncorrelated, the ideal-observer SNR takes on a particularly simple form,

$$SNR_I^2 = \Delta s^t \left(diag\{Hf\} \right)^{-1} \Delta s, \qquad (4)$$

where Δs is the signal vector, $diag\{ \}$ is a diagonal matrix and Hf is the noise free projection data [31].

C. Data Acquisition and Processing

As discussed above, computing the ideal-observer SNR for a given imaging geometry and processing approach requires two ensembles of images: one set consisting of images of the signal and background together and one set consisting of images of the background alone. In order to preserve the flexibility to compute the ideal-observer SNR for lesions of varying concentration, we acquired high-count projection images of the signal alone (i.e., in a cold background), which could be scaled as desired and added to the ensemble of background projections alone to produce an ensemble of signal-plus-background projections. These were then either analyzed directly for the planar geometry or processed and reconstructed for the tomographic geometries. Of course, for linear techniques, we could have computed $|\Delta S_{out}(\upsilon)|^2$ from the images of the signal alone, because the

difference between the ensemble averages in this case is simply equal to the ensemble average of the differences—the signal we added to the background ensemble. However, this equivalence does not generally hold for non-linear techniques such as the adaptive smoothing being investigated.

Our phantom consisted of an 14-cm diameter, 800 cc cylinder with a 1-cm outer diameter spherical lesion insert. This lesion size is representative of the smallest currently detected in scintimammography. For each geometry we acquired 20 images of the background alone, for which we put 3.7 mCi of activity in the phantom and imaged for 1 minute total for each conventional or dedicated SPECT acquisition and 30 seconds for each planar view. These combinations of activity and imaging time were chosen to produce clinically realistic count levels using the following reasoning. As with Wang et al., we began with the assumption that 1% of a typical 25 mCi clinical dose of Tc-99m-sestamibi is taken up by the myocardium. Using the volume of the myocardium in the Data Spectrum Corporation cardiac insert as a guide, along with the assumption that soft tissue will have a 1:15 concentration relative to the myocardium allowed us to determine the expected concentration of activity in healthy breast tissue. We wished to compare detectability in these three geometries given the same total imaging time, and assumed that typical clinical imaging times would be 30 minutes per SPECT study and 15 minutes per planar view. Thus we were comparing the three geometries for equal total imaging times given that two-view planar studies are common. Given this, we scaled up the calculated concentration by a factor of 30 and scaled down the imaging times by the same factor to allow for more rapid data acquisition. All imaging times were adjusted to compensate for the decay of the activity. To image the lesion we filled it with 7.6 mCi of Tc-99m, placed it in the cylinder now filled with cold (zero-activity) water, and imaged for 30 minutes in conventional and dedicated SPECT and 20 minutes for the planar view. This combination of activity and imaging time was chosen simply to provide high-count, low noise images of the signal, which could be scaled appropriately and added to the background images.

The dedicated SPECT images were acquired by placing the phantom at the center of rotation of a Picker XP2000 two-headed SPECT system with the heads rotating at the minimum radius of rotation (9.0 cm). In this configuration the heads were within 2.0 cm of the walls of the phantom. The breast phantom was not attached to an anthropomorphic torso phantom because Wang et al. showed that with proper shielding the contribution of scatter from the torso can be made negligible. We acquired 120 views over 360° with each head acquiring to a 128x128 matrix (pixel size=4.67 mm). We used a low-energy, ultra-high resolution collimator. The conventional SPECT images were also acquired in the absence of an anthropomorphic torso phantom, although the radius of rotation (25 cm) and the placement of the breast phantom (17 cm off-center) were determined with the torso phantom in place. The reason for this curious arrangement was to isolate the effect of the large radius of rotation on lesion detectability, without the additional degradations caused by attenuation or scatter in the torso. The ideal-observer SNR results for this arrangement will thus represent

an upper bound on the true detectability. This arrangement also facilitates computation of the Wiener spectrum, which requires images of a stationary, uniform background that would have been difficult to achieve in the presence of the highly non-uniform attenuation caused by the thorax. In other respects, the conventional SPECT images used the same acquisition parameters as the dedicated SPECT. Finally, the planar views were acquired with the phantom flush against one head, which acquired on a 128x128 matrix with a magnification factor of 2.0 (pixel size=2.33 mm).

For the tomographic geometries, ideal-observer SNRs for the sinograms were calculated using equation (4), while SNRs for reconstructed images were computed by the following procedure:

1. The signal projections were scaled to simulate a desired tumor-background concentration ratio (6:1 in this case) and added to each of the 20 sets of background projections.

2. The slice through the center of the lesion was selected and the 20 corresponding signal-plus-background sinograms were reconstructed by filtered backprojection using ramp and Hanning filters with various cutoff frequencies (0.4, 0.6, 0.8, and 1.0 times the Nyquist frequency). The sinograms were also processed using the effective 2D smoothing procedure described above and reconstructed by filtered backprojection.

3. The 20 corresponding sinograms of background alone were processed in the same way.

4. An average signal image was determined by subtracting the average of the 20 background alone reconstructions from the average of the 20 signal-plus-background reconstructions. The signal power spectrum was computed by squaring the Fourier transform of this image.

5. While SPECT images are not stationary in general, the attenuated projections of a uniform cylinder of this diameter are quite flat over a broad central region, and thus one might reasonably expect the reconstructed images of this cylinder to be locally stationary near their center, precisely where the lesion is expected to lie [32]. The "local" Wiener spectrum in this region was computed from the 20 images of background alone by subtracting the average background image from each of the individual background images, resulting in 20 noise images. Each such image was multiplied by a circularly symmetric window of the form:

$$w(\mathbf{r}) = 1 \quad \text{for } |\mathbf{r}| \le 0.9R,$$

$$w(\mathbf{r}) = 0.5 \times (1 + \cos(\pi(|\mathbf{r}| - 0.9R) / 0.2R),$$
$$\text{for } 0.9R < |\mathbf{r}| < 1.1R, \text{ and} \tag{5}$$

$$w(\mathbf{r}) = 0 \quad \text{for } |\mathbf{r}| \ge 1.1R,$$

(with appropriate shifting for the off-center cylinder in the conventional geometry), where R is the radius of the circular region over which the noise is expected to be stationary (chosen to 6 pixels) and $\mathbf{r}$ the radial position in the image. The result was centered and zero-padded to 128x128. The noise power spectrum of each of the 20 images was computed by taking the square of the Fourier transform of the resulting image. The 20 noise power spectra were then averaged and scaled so that the volume under the Wiener spectrum equaled the average variance in the circle of radius R [33-35].

6. The ideal observer SNR was then determined by summing the quotient of the calculated signal spectrum and Wiener spectrum.

For planar views, the procedure was essentially the same, without the reconstruction step but including the application of the effective 2D smoothing. Signal spectra were determined from the difference between the ensemble averages of signal-plus-background images and background images and Wiener spectra from the background images alone, using the same rolled-off cylindrical window to isolate a reasonably uniform region and to minimize truncation effects. As a consistency check, equation (4) was also used to calculate the SNR of the unsmoothed planar dataset.

Finally it should be noted that in using the ideal-observer framework at all it is implicitly being assumed that the data satisfy the assumptions discussed in section II.B: that the system is linear and that the noise in the planar or reconstructed images is additive, Gaussian, zero-mean, stationary, and independent of the presence or absence of the signal. Given the reasonably high count levels (~10-15/pixel) the fact that the signal is relatively small and low contrast, and the discussion of stationarity in point 5 above, the assumptions about the noise seem reasonable. The requirement of linearity seemingly undermines the use of the framework to analyze images that have been processed by adaptive, effective multi-dimensional smoothing. However, what is truly required for equation (3) to be meaningful is not linearity in the face of any possible input but more specifically that the system transfer function be the same whether the particular signal of interest is present or absent from the particular background of interest. Again, because the signal in question is relatively small and low contrast, it should not greatly affect the noise properties of the projection images and thus the effective multi-dimensional smoothing algorithm should yield a similar effective system transfer function whether or not the signal is present.

III. RESULTS

The unsmoothed and smoothed planar views are shown in Figure 1. The ideal-observer SNR for the planar views was found to be 6.2 without processing (use of equations (3) and (4) yielded the same result) and essentially the same, 6.4, after adaptive, effective multi-dimensional smoothing
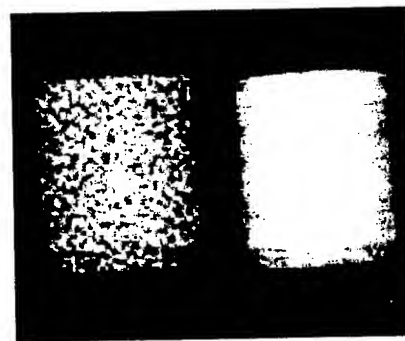

Figure 1: Planar images of a cylindrical phantom containing a 1-cm, 6:1 lesion. The image on the left is unprocessed; the image on the right has undergone adaptive, effective 2D smoothing. The image on the left corresponds to a calculated SNR of 6.2, the image on the right to an SNR of 6.4.

Table 1
Ideal Observer SNRs for conventional and dedicated SPECT

Processing Method	Dedicated Breast SPECT	Conventional SPECT[†]
Sinogram	10.7	6.7
Hanning (cutoff=0.4)	6.2	2.4
Hanning (cutoff=0.6)	8.0	3.1
Hanning (cutoff=0.8)	9.4	3.5
Hanning (cutoff=1.0)	9.9	3.7
Ramp (cutoff=0.4)	8.1	3.3
Ramp (cutoff=0.6)	9.7	3.7
Ramp (cutoff=0.8)	9.9	3.6
Ramp (cutoff=1.0)	9.3	3.7
Eff. 2D smoothing	10.6	5.5

[†]The conventional SPECT SNRs represent an upper bound on detectability (see section II.C for details).

The ideal-observer SNRs for the detection of a lesion with a 6:1 lesion-background concentration ratio are listed in Table 1 for the two different tomographic geometries.

It has been shown experimentally that the minimum SNR necessary for a human observer to be able to detect reliably a signal in a noisy background is 5.0 [36]. Regardless of the processing approach, the conventional SPECT and planar geometries yield SNRs that are below or only just above this threshold. The results above thus confirm quantitatively the findings of Wang *et al.* that a 6:1 lesion of this size is difficult or impossible to detect using planar or conventional SPECT geometries, but quite reliably detectable using a dedicated geometry. The results also indicate that effective two-dimensional smoothing provides improvement in SNR over other filtering approaches. Images corresponding to the two different geometries for selected processing methods are shown in Figure 2. These confirm visually the conclusions just stated: the 6:1 lesion is quite visible in all of the dedicated reconstructions, while it is effectively undetectable in the conventional reconstructions. The 6:1 lesion is rather difficult to discern in both the unprocessed and processed planar views depicted in Figure 1.

IV. DISCUSSION AND CONCLUSIONS

The SNR values given in Table 1 suggest that a dedicated SPECT geometry would lead to improved detectability for clinically typical lesions over the planar and conventional SPECT geometries. Recall that in the presence of attenuation and scatter from the torso we would expect the difference between the dedicated and conventional SPECT geometries to be even greater than it is here. The success of the dedicated geometry can be attributed to the fact that it combines the advantages of the other two approaches: because of its small radius of rotation, it offers good sensitivity and resolution comparable to that of a planar view acquired with the

scintillation camera flush against the phantom, while it offers the improved contrast offered by a tomographic system's ability to separate lesion activity from overlying and underlying activity.
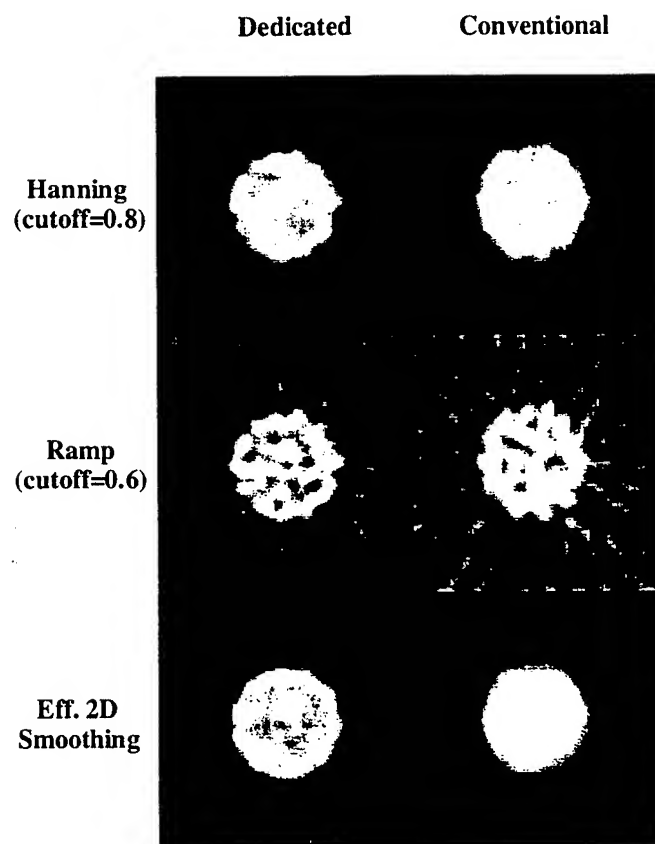

Figure 2: Reconstructed slices of a cylindrical phantom containing a 1-cm, 6:1 lesion for two different tomographic geometries.

The SNR values also support the hypothesis that adaptive, effective multi-dimensional smoothing may improve lesion detectability relative to the other kinds of filtering used in tomographic reconstruction. It is a fact that processing or even image reconstruction can never improve the ideal-observer SNR over that found in the raw projection data. However, these operations can certainly *diminish* the SNR if they are in some way singular and if the signal vector has components in the null space. We observe that all the reconstructed image SNRs are lower than those of their corresponding sinograms. The differences among those reconstructed image SNRs give some clue as to how much of the information contained in those projections is preserved in the reconstructed image. It would seem that the use of the effective multi-dimensional smoothing prior to reconstruction allows more of the information in the projections to persist in the reconstructed images.

Finally, it should be noted that in the case of fully linear imaging process, the calculated ideal-observer SNR should be directly proportional to the lesion concentration and thus to the tumor-background concentration ratio. This allows us to express the SNRs as a function of tumor-background

concentration ratio, C, yielding for instance an SNR of $1.7*C$ for the dedicated SPECT geometry using a Hanning filter (cutoff=1.0), an SNR of $0.6*C$ for the conventional SPECT geometry using the same filter, and an SNR of $1.0*C$ for the unprocessed planar geometry. Assuming that an SNR of 5.0 corresponds to the threshold of detectability, this allows us to conclude that for the specified imaging times and lesion size, the minimum tumor-background ratios required for detectability are approximately 2.9-to-1 for the dedicated geometry, 8.3-to-1 for the conventional geometry, and 5.0-to-1 for the planar geometry using the specified filters.

Throughout this study we have been comparing the three geometries on the assumption of equal total imaging times (30 minutes) and implicitly assuming that in the planar geometry we were examining the one of the two 15-minute views in which the lesion appeared most prominently. It is not necessarily possible to identify this view *a priori*, but assuming it were we can estimate the SNR corresponding to a 30-minute acquisition at that single view. Because SNRs for linear methods should in principle scale as the square root of the imaging time, the SNR would be approximately $6.2 * \sqrt{2} = 8.8$, or comparable to the dedicated SPECT SNR.

The dependence of the ideal-observer SNR on lesion size remains a subject for future investigation.

V. ACKNOWLEDGMENTS

VI. REFERENCES

[1] Cancer Facts & Figures, Atlanta: American Cancer Society, 1996.

[2] I. Andersson, "Mammographic screening and mortality from breast cancer: Malmö mammographic screening trial," *Br. J. Med.*, vol. 297, pp. 943-948, 1988.

[3] J. Frisell, G. Eklund, L. Hellström, E. Lidbrink, L.E. Rutqvist and A. Somell, "Randomized study of mammographic screening—preliminary report on mortality in the Stockholm trial," *Breast Cancer Res. Treat.*, vol. 18, pp. 49-56, 1991.

[4] A.B. Miller, C.J. Baines, T. To and C. Wall, "Canadian national breast screening study," *Can. Med. Assoc. J.*, vol. 147, pp. 1459-1476, 1992.

[5] D.B. Kopans, "The positive predictive value of mammography," *AJR*, vol. 158, pp. 521-526, 1992.

[6] E. Bombardieri, F. Crippa, L. Maffioli and M. Greco, "Nuclear medicine techniques for the study of breast cancer," *Eur. J. Nucl. Med.*, vol. 24, pp. 809-824, 1997.

[7] A.D. Waxman, "The role of Tc-99m Methoxy-isobutylisonitrile in imaging breast cancer," *Seminars in Nuclear Medicine*, vol. 27, pp. 40-54, 1997.

[8] C.H. Kao, S.J. Wang and T.J. Liu, "The use of technetium-99m methoxyisobutylisonitrile breast scintigraphy to evaluate palpable breast masses," *Eur. J. Nucl. Med.*, vol. 21, pp. 432-436, 1994.

[9] I. Khalkhali, J. Cutrone, I. Mena, L. Diggles, R. Venegas, H. Vargas, B. Jackson and S. Klein, "Technetium-99m-sestamibi Scintimammography of Breast Lesions: Clinical and Pathological Follow-up," *J. Nucl. Med.*, vol. 36, pp. 1784-1789, 1995.

[10] I. Khalkhali, I. Mena and L. Diggles, "Review of imaging techniques for the diagnosis of breast canacer: a new role of prone scintimammography using technetium-99m-sestamibi," *Eur. J. Nucl. Med.*, vol., 21, pp. 357-362, 1994.

[11] I. Khalkhali, J. Cutrone, I. Mena, L. Diggles, R.J. Venegas, H.I. Vargas, B.L. Jackson, S. Khalkhali, J.F. Moss and S.R. Klein, "Scintimammography: the complementary role of Tc-99m sestamibi prone breast imaging for the diagnosis of breast carcinoma," *Radiology*, vol. 196, pp. 421-426, 1995.

[12] J. Villanueva-Meyer, M.H. Leonard Jr., E. Briscoe, F. Cesani, S.A. Ali, S. Rhoden, M. Hove and D. Cowan, "Mammoscintigraphy with Technetium-99m-sestamibi in suspected breast cancer," *J. Nucl. Med.*, vol. 37, pp. 926-930, 1996.

[13] R. Taillefer, A. Robidoux, R. Lambert, S. Turpin and Jean Laperrière, "Technetium-99m-sestamibi prone scintimammography to detect primary breast cancer and axillary lymph node involvement," *J. Nucl. Med.*, vol. 36, pp. 1758-1765.

[14] H. Palmedo, A. Schomburg, F. Grünwald, P. Mallman, D. Krebs and H.-J. Biersack, "Technetium-99m-MIBI scintimammography for suspicious breast lesions," *J. Nucl. Med.*, vol. 37, pp. 626-630, 1996.

[15] H. Palmedo, F. Grünwald, H. Bender, A. Schomburg, P. Mallmann, D. Krebs and H.J. Biersack, "Scintimammography with technetium-99m methoxy-isobutylisonitrile: comparison with mammography and magnetic resonance imaging," *Eur. J. Nucl. Med*, vol. 23, pp. 940-946, 1996.

[16] F. Scopinaro, M. Ierardi, L.M. Porfiri, N.S. Tiberio, G. De Vincentis, S. Mezi, P. Cannas, T. Gigliotti and L. Marzetti, "Tc-99m-MIBI prone scintimammography in patients with high and intermediate risk mammography," *Anticancer Research*, vol. 17, pp. 1635-1638, 1997.

[17] F. Scopinaro, O. Schillaci, W. Ussof, K. Nordling, R. Capoferro, G. De Vincentis, R. Danieli, M. Ierardi, V. Picardi, R. Tavolaro and A.C. Colella, "A three-center study on the diagnostic accuracy of 99mTc-MIBI scintimammography," *Anticancer Reaserch*, vol. 17, pp. 1631-1634, 1997.

[18] A. Maurier, D.F. Caroline, F.J. Jadali, T.A. Manzone, W.P. Maier, F.C. Au, and S.F. Schnall, "Limitations of caraniocaudal thallium-201 and technetium-99m-sestamibi mammoscintigraphy," *J. Nucl. Med.*, vol. 36, pp. 1696-1700, 1995.

[19] J. Maublant, M. de Latour, D. Mestas, A. Clemenson, S. Charrier, V. Feillel, G. Le Bouedec, P. Kaufmann, J. Dauplat and A. Veyre, "Technetium-99m-sestamibi uptake in breast tumor and associated lymph nodes," *J. Nucl. Med.*, vol. 37, pp. 922-925, 1996.

[20] L. Diggles, I. Mena and I. Khalkhali, "Technical aspects of prone dependent-breast scintimammography," *J. Nucl. Med. Tech.*, vol. 22, pp. 165-170, 1994.

[21] M.A. Nathan, J.E. Seabold, T. Barloon, J. Quesenberry, D.L. Bushnell, R.A. Robinson and D.C. Young, "Planar vs. SPECT tc-99m MIBI evaluation of suspicious breast lesions on mammograohy: histologic correlation," *J. Nucl. Med.*, vol. 35, p. 229P, 1994.

[22] N. Nagaraj, A. Waxman, G. Ashok, S. Khan, L. Memsic, J. Yadegar and E. Phillips, "Comparison of SPECT and planar Tc-99m sestamibi (MIBI) imaging in patients with carcinoma of the breast," *J. Nucl. Med.*, vol. 35, p. 299P, 1994.

[23] R. Tiling, M. Pechmann, H. Sommer, R. Moser, K. Tatsch and K. Khan, "Does SPECT improve the diagnostic accuracy of planar scintimammography with sestamibi?" *J. Nucl. Med.*, vol. 23, pp. 97-102, 1982.

[24] H. Wang, C. Scarfone, K.L. Greer, R.E. Coleman and R.J. Jaszczak, "Prone breast tumor imaging using vertical axis-of-rotation SPECT systems: an intital study," *IEEE Trans. Nucl. Sci.*, vol. 44, pp. 1271-1276, 1997.

[25] A. Williams and K. Burrage, "Surface fitting using GCV smoothing splines on supercomputers," *CD-ROM proceedings of the 1995 ACM/IEEE conference on supercomputing*, 1995. The proceedings are also available on the World Wide Web at *http://www.supercomp.org/sc95/proceedings.*

[26] X. Pan, "A general approach for multidimensional smoothing," *Med. Phys.* (in press), 1998.

[27] P.J. Green and B.W. Silverman, *Nonparametric Regression and Generalized Linear Models*, London: Chapman Hall, pp. 11-27, 1994.

[28] J.A. Fessler, "Penalized weighted least-squares image reconstruction for positron emission tomography," *IEEE Trans. Med. Imag.*, vol. 13, pp. 290-300, 1994.

[29] D.M. Bates, M.J. Lindstrom, G. Wahba and B.S. Yandell, "GCVPACK routines for generalized cross validation," *Commun. Statist. Simul. Comput.*, vol. 16, pp. 263-297, 1987.

[30] "Medical imaging—the assessment of image quality," *ICRU Report 54*, 1996. Available from International Commission on Radiation Units and Measurements, 7910 Woodmont Ave., Bethesda, MD, 20814.

[31] J. Yao and H.H. Barrett, "Predicting human performance by a channelized Hotelling observer model," *Proc. SPIE*, vol. 1768, pp. 161-168, 1992.

[32] S.C. Moore, M.F. Kijewski, S.P. Müller and B.L. Holman, "SPECT image noise power: effects of nonstationary projection noise and attenuation compensation," *JNM*, vol. 29, pp. 1704-1709, 1988.

[33] S.J. Riederer, N.J. Pelc and D.A. Chesler, "The noise power spectrum in computed x-ray tomography," *Phys. Med. Biol.*, vol. 23, pp. 446-454, 1978.

[34] K. Hanson, "Detectability in computed tomographic images," *Med. Phys.*, vol. 6, pp. 441-451, 1979.

[35] M.F. Kijewski and P.F. Judy, "The noise power spectrum of CT Images," *Phys. Med. Biol.*, vol. 32, pp. 565-575, 1987.

[36] A. Rose, *Vision: Human and Electronic*, New York: Plenum Press, 1973.

2224

Spline-Based Inverse Radon Transform in Two and Three Dimensions

P.J. La Rivière, *Student Member, IEEE*, and X. Pan, *Member, IEEE*

Department of Radiology, The University of Chicago, Chicago, IL, 60637

Abstract

While the exact inverse Radon transform is a continuous integral equation, the discrete nature of the data output by tomographic imaging systems generally demands that images be reconstructed using a discrete approximation to the transform. However, by fitting an analytic function to the projection data prior to reconstruction, one can avoid such approximations and preserve and exploit the continuous nature of the inverse transform.

We present methods for the evaluation of the inverse Radon transform in two and three dimensions in which cubic spline functions are fit to the projection data, allowing the integrals that represent the filtration of the sinogram to be carried out in closed form and also eliminating the need for interpolation upon backprojection. Moreover, in the presence of noise, the algorithm can be used to reconstruct directly from the coefficients of *smoothing* splines, which are the minimizers of a popular curve-fitting measure. We find that the 2D and 3D direct-spline algorithms have superior resolution to their 2D and 3D FBP counterparts, albeit with higher noise levels, and that they have slightly lower ideal-observer signal-to-noise ratios for the detection of a 1-cm, spherical lesion with a 6:1 lesion-background concentration ratio.

I. INTRODUCTION

The inverse Radon transform provides the mathematical foundation for tomographic imaging, which involves reconstructing images of distributions of anatomical or physiological properties from projections of those distributions. In computed tomography (CT) [1], for instance, the property being imaged is the linear photon attenuation coefficient of tissue at various points in the body, while in positron emission tomography (PET) [2] or in single-photon emission computed tomography (SPECT) [3] it is the concentration of injected radiopharmaceuticals at various points in the body. We will denote any such distribution in two dimensions by $f(x,y)$ and label each projection through it by the pair $\{\xi,\varphi\}$, where φ specifies the *projection angle* and ξ the *projection distance*. The value of such a projection is given by the line integral of the distribution along the line specified by $\{\xi,\varphi\}$ and will be denoted by $p(\xi,\varphi)$. Similarly, we will denote a three-dimensional distribution by $f(x,y,z)$ and label each projection through it by the triplet $\{\xi,\theta,\varphi\}$, where θ and φ specify the orientation of the plane and ξ the distance of the plane to the origin of the coordinates. The value of such a projection is given by the *planar* integral of the distribution over the plane specified by $\{\xi,\theta,\varphi\}$ and will be denoted by $p(\xi,\theta,\varphi)$. The functions $p(\xi,\varphi)$ and $p(\xi,\theta,\varphi)$ are known as *sinograms* because in the two-dimensional case a point

distribution in $\{x,y\}$ space maps to a sinusoid in $\{\xi,\varphi\}$ space.

The Radon transform is a continuous, integral transform that relates the sinogram $p(\xi,\varphi)$ to $f(x,y)$ in two dimensions and the sinogram $p(\xi,\theta,\varphi)$ to $f(x,y,z)$ in three dimensions [4,5,6]. Inverting the Radon transform exactly to recover a distribution requires continuous, noise-free knowledge of the distribution's sinogram, which entails having an infinite set of perfect projection measurements. In practice, of course, one can only collect projection data in the two-dimensional case for a finite number of projection distances ξ_i (we call these discrete samples *projection bins*) at a finite number of projection angles φ_j, and the measurements are invariably contaminated with noise. In the three-dimensional case, the planar-integral projection data cannot generally be measured directly and must instead be generated from line-integral projection data; it is, however, still only generated for a finite number of projection bins ξ_i and projection angles φ_j and θ_k.

Because the sinogram $p(\xi,\varphi)$ in two dimensions (or $p(\xi,\theta,\varphi)$ in three dimensions) is known only on a finite grid of ξ_i and φ_j (or ξ_i, φ_j, and θ_k), we cannot invert the Radon transform exactly to recover the distribution $f(x,y)$ (or $f(x,y,z)$); we must instead turn to a discrete approximation of the inverse. For instance, one way of implementing the most popular two-dimensional Radon inversion algorithm—filtered backprojection (FBP) [7,8]—begins with a discrete filtration of the sinogram. The filtered samples of ξ_i for each projection angle φ_j are then backprojected onto the image grid and the resulting images summed to give a final reconstructed image. One way to view the backprojection step is to imagine casting a perpendicular ray from each image pixel to each projection angle in turn, summing the sinogram values picked up at each angle to obtain the final pixel value. In this view, the difficulty lies in determining what value to pick up from each projection angle, for in general the perpendicular line will not fall directly in the center of a projection bin. In the simplest schemes, one simply picks up the value of the bin the pixel projects onto, while in a more complicated approach one might perform a linear interpolation of the values in the two nearest projection bins. In the most sophisticated schemes, one uses a weighted average of the values in a slightly larger neighborhood. A similar procedure can be used to implement three-dimensional FBP [9].

The discreteness of the sinogram thus dictates discreteness in both the filtration and backprojection steps of the algorithm. If, however, one had a continuous, analytic expression for the sinogram at each projection angle—if the sinogram were a set of known one-dimensional functions of ξ—it might be possible to implement the filtration and backprojection steps in a continuous manner. The filtration could be performed analytically, and the resulting filtered

projections would be continuous functions which, in the pixel-traversal view of backprojection described above, could be sampled wherever a projection might strike without any need to interpolate. Naturally, such an analytic, continuous expression for the sinogram cannot be obtained directly from any tomographic imaging system, but must rather be obtained by fitting an analytic expression to the discretely sampled data. Not every class of function that could be fit to the data would allow the filtration of the projections to be calculated in closed form, but Wahba [10] has shown that one class that does allow such a solution are the cubic splines, piecewise third-order polynomials that are continuous up to and including the second derivative at the connection points between pieces. This is the class of fitting functions we investigate in this paper, introducing Wahba's results (with some corrections and simplifications), and extending the treatment to the three-dimensional Radon transform.

This method offers a certain conceptual appeal, as well as the advantage of directness when one wishes to smooth noisy projection data by fitting curves that minimize the popular penalized least-squares measure [11]. As it turns out, the minimizers of this measure are the natural cubic splines mentioned above and thus reconstruction can proceed directly from the coefficients of these splines in this case.

II. METHODS

A. Inverse 2D Radon transform in coordinate space

The essential problem in two-dimensional tomography is the reconstruction of a distribution $f(x,y)$ (or $f(r,\theta)$ in polar coordinates) from knowledge of the discrete sinogram $p(\xi_i, \varphi_j)$, where $i = -N,...,N$ and $j = 1,...,M$. This convention for the index i, particularly the choice of an odd number of projection bins, will simplify the mathematical expressions to be derived below, but the proposed technique is applicable to geometries with an even number of bins as well. We assume in the present paper that we have a parallel-beam geometry.

Perhaps the most familiar way of expressing the inverse of the two-dimensional Radon transform is in terms of the frequency-space representation of the continuous sinogram:

$$f(r,\theta) = \int_0^\pi \int_{-\infty}^\infty |v| P(v,\varphi) e^{j2\pi v \xi} dv \, d\varphi, \qquad (1)$$

where v is the spatial-frequency variable corresponding to ξ, $P(v,\varphi)$ is the Fourier transform of the sinogram $p(\xi,\varphi)$ with respect to the variable ξ, and j is the imaginary number $\sqrt{-1}$. This expression provides the theoretical basis for FBP, as $|v|$ is just the familiar ramp filter. This expression may be written in coordinate space as

$$f(r,\theta) = \frac{1}{2\pi^2} \int_0^\pi J_{r,\theta}(\varphi) d\varphi, \qquad (2)$$

where

$$J_{r,\theta}(\varphi) = \left\{ \lim_{\varepsilon \to 0} \left[\int_{-\infty}^{\xi'-\varepsilon} \frac{p'(\xi,\varphi)}{\xi'-\xi} d\xi + \int_{\xi'+\varepsilon}^{\infty} \frac{p'(\xi,\varphi)}{\xi'-\xi} d\xi \right] \right\}, \qquad (3)$$

and in which $\xi' = r\cos(\theta - \varphi)$ and $p'(\xi,\varphi)$ is the first derivative of the sinogram $p(\xi,\varphi)$ with respect to ξ [12,4]. Taking the limit as $\varepsilon \to 0$ of the sum of these two integrals allows us to avoid integrating over the singularity at $\xi' = \xi$. In general equations (2) and (3) are less useful than equation (1) because the data collected in PET, SPECT, or CT constitute samples of the sinogram $p(\xi,\varphi)$ itself, and do not provide any direct information about $p'(\xi,\varphi)$. However, by fitting an analytic, differentiable function of ξ to the projection data at each angle, we could obtain an expression for $p'(\xi,\varphi)$. If $p'(\xi,\varphi)$ had an auspicious functional form, we would then be able to solve the integrals in equation (3) in closed form.

B. Interpolating and smoothing splines

To obtain an expression for the sinogram that is analytic and differentiable with respect to the variable ξ, we fit a function to the projection data at each angle. That is, for each angle φ_j, we fit a one-dimensional function of ξ, $\hat{p}_{\varphi_j}(\xi)$, to the sinogram values $p_{\varphi_j}(\xi_i)$ measured on the $2N+1$ abscissas ξ_i $(i = -N,...,N)$. If the data is noiseless, it is desirable to use a function that passes through the points $p_{\varphi_j}(\xi_i)$, which we call an *interpolating* curve, while if the data is noisy a *smoothing* curve may be more appropriate. One fitting framework that can handle both of these situations is known as penalized least-squares [11], in which the function $\hat{p}_{\varphi_j}(\xi)$ is chosen to be the minimizer of the functional

$$F\left[\hat{p}_{\varphi_j}(\xi)\right] = \sum_{i=-N}^N \left[p_{\varphi_j}(\xi_i) - \hat{p}_{\varphi_j}(\xi_i) \right]^2 + \lambda \int_a^b \left(\hat{p}_{\varphi_j}''(\xi) \right)^2 d\xi, \qquad (4)$$

where a and b are the endpoints of the interval on which the curve $\hat{p}_{\varphi_j}(\xi)$ is to be defined. The first term in equation (4) is the familiar squared-error measure, while the second is a measure of the smoothness of the fit curve. The parameter λ thus mediates the tradeoff between the competing goals of achieving a good fit to the data and maintaining a smooth curve. By choosing λ to be zero, we eliminate the smoothness constraint and ensure that the minimizing curve will be an interpolant to the data; if λ grows large, the smoothness constraint dominates and the curve approaches a linear fit to the data. For intermediate values of λ the minimizing curve balances the goodness-of-fit and smoothness constraints. The parameter λ can be chosen *a priori* [13] or it can be determined from the statistics of the data using an automatic procedure such as generalized cross-validation (GCV) [14].

The minimizers of this functional F belong to the class of functions known as natural cubic splines [13]. These are piecewise polynomial curves that join at the abscissa values ξ_i, where they are continuous up to and including the second derivative. They can be represented as

$$\hat{p}_{\varphi_j}(\xi) = a_i + b_i \xi + c_i \xi^2/2 + d_i \xi^3/3, \qquad \xi \in [\xi_i, \xi_{i+1}], \quad (5)$$

where a_i, b_i, c_i, and d_i are constants that fully specify the spline curve on the interval $[\xi_i, \xi_{i+1}]$. Of interest to the Radon inversion problem is the fact that we can approximate the first derivative of the sinogram for fixed angle φ_j by

$$p'(\xi, \varphi_j) \cong \hat{p}_{\varphi_j}'(\xi) = b_i + c_i \xi + d_i \xi^2, \qquad \xi \in [\xi_i, \xi_{i+1}]. \quad (6)$$

C. Spline-based inverse of the 2D Radon transform

Given this analytic expression for the derivative of the sinogram, we can proceed with the inversion of the two-dimensional Radon transform in equation (2). While the sinogram now has a continuous representation in the variable ξ, it is still discrete in the angular variable. Assuming that the M angular samples are equally spaced over 180° or 360° (the result is the same in either case), the integral in equation (2) can be approximated by the sum

$$f(r,\theta) \cong \frac{1}{2\pi M}\sum_{j=1}^{M} J_{r,\theta}(\varphi_j), \qquad (7)$$

where $J_{r,\theta}(\varphi_j)$ is given by equation (3). For a given coordinate (r,θ) in image space, and for a given projection angle φ_j, $\xi' = r\cos(\theta - \varphi_j)$. We label the projection bin that ξ' falls in by m, that is, $\xi' \in [\xi_m, \xi_{m+1}]$. Using the expression for $p'(\xi,\varphi)$ given by equation (6), $J_{r,\theta}(\varphi_j)$ can be expressed as

$$J_{r,\theta}(\varphi_j) = \sum_{\substack{i=-N \\ i\neq m}}^{N-1}\int_{\xi_i}^{\xi_{i+1}} \frac{b_i + c_i\xi + d_i\xi^2}{\xi'-\xi}d\xi$$
$$+\lim_{\varepsilon\to 0}\left[\int_{\xi_m}^{\xi'-\varepsilon} \frac{b_m + c_m\xi + d_m\xi^2}{\xi'-\xi}d\xi + \int_{\xi'+\varepsilon}^{\xi_{m+1}} \frac{b_m + c_m\xi + d_m\xi^2}{\xi'-\xi}d\xi\right], \qquad (8)$$

where the integrals of equation (3) are now expressed as sums of integrals over the subintervals between the original abscissa points, with appropriate accommodation for the singularity at ξ'. These integrals can be solved in closed form [10], and the resulting expression contains some potentially unstable terms. However, by combining these terms in a particular way and invoking spline identities, a stable form can be derived. The details are given in appendix A, where the final expression for $J_{r,\theta}(\varphi_j)$ is shown to be

$$J_{r,\theta}(\varphi_j) = T + \sum_{\substack{i=-N \\ i\neq m-1,m,m+1}}^{N-1}\left(b_i + c_i\xi' + d_i\xi'^2\right)\ln\left(\frac{\xi'-\xi_i}{\xi'-\xi_{i+1}}\right)$$
$$-\sum_{i=-N}^{N-1}(c_i + 2d_i\xi')(\xi_{i+1}-\xi_i) + \frac{1}{2}\sum_{i=-N}^{N-1}d_i\left[(\xi'-\xi_i)^2 - (\xi'-\xi_{i+1})^2\right] \qquad (9)$$

where T is given by equation (A.4) of appendix A.

D. The 3D Radon transform in coordinate space

The essential problem in three-dimensional computed tomography is the reconstruction of a distribution $f(x,y,z)$ from knowledge of the discrete planar-integral sinogram $p(\xi_i,\theta_k,\varphi_j)$, where $i = -N,...,N$, $j = 1,...,M_\varphi$, and $k = 1,...,M_\theta$ [15]. In general, these planar integrals are not measured directly by tomographic imaging systems, but must rather be calculated by "rebinning" the line integrals that are measured directly [9].

The inverse three-dimensional Radon transform has a form similar to the two-dimensional case, with a few differences that greatly simplify the task of evaluating it numerically. Specifically,

$$f(x,y,z) = \frac{-1}{4\pi^2}\int_0^\pi\int_0^\pi p''(\xi',\theta,\varphi)\sin\theta\,d\varphi\,d\theta, \qquad (10)$$

where

$$\xi' = x\sin\theta\cos\varphi + y\sin\theta\sin\varphi + z\cos\theta, \qquad (11)$$

and $p''(\xi',\theta,\varphi)$ is the second derivative of the three-dimensional sinogram with respect to ξ'. This expression differs in two principal ways from the expression for the two-dimensional inverse Radon transform given by equations (2) and (3). First, it now involves the second derivative of the sinogram with respect to ξ' rather than the first derivative and second, the convolution in ξ has disappeared. This reflects the fact that an inverse Radon transform of odd degree can be calculated using purely *local* information—the value of the image at a point (x,y,z) can be determined solely from information at points in the sinogram space that (x,y,z) projects onto, rather than from a convolution integral over all points in sinogram space as in the even-dimensional case [5,16].

E. Spline-based inverse of the 3D Radon transform

As in the two-dimensional case, we wish to fit an analytic function of ξ to each sequence of ξ_i labeled by a distinct pair $\{\varphi_j,\theta_k\}$. We do so using the spline formalism described above and obtain

$$\hat{p}_{\varphi_j,\theta_k}(\xi) = a_i + b_i\xi + c_i\xi^2/2 + d_i\xi^3/3, \quad \xi \in [\xi_i, \xi_{i+1}]. \quad (12)$$

Consequently the second derivative of the sinogram is approximately

$$p''(\xi,\theta_k,\varphi_j) \cong \hat{p}''_{\varphi_j,\theta_k}(\xi) = c_i + 2d_i\xi, \quad \xi \in [\xi_i,\xi_{i+1}]. \quad (13)$$

Now as in the two-dimensional case, the discreteness of the angular samples means that the two integrals give way to sums and we write

$$f(x,y,z) \cong -\frac{1}{4}\frac{1}{M_\theta}\frac{1}{M_\varphi}\sum_{j=1}^{M_\varphi}\sum_{k=1}^{M_\theta} J_{x,y,z}(\theta_k,\varphi_j)\sin\theta_k, \qquad (14)$$

where

$$J_{x,y,z}(\theta_k,\varphi_j) = c_i + 2d_i\xi', \quad \xi' \in [\xi_i,\xi_{i+1}], \qquad (15)$$

and

$$\xi' = x\sin\theta_k\cos\varphi_j + y\sin\theta_k\sin\varphi_j + z\cos\theta_k. \qquad (16)$$

The simplicity of the three-dimensional inversion is now apparent, for the functions J can be evaluated in a straightforward manner whereas in the two-dimensional case evaluation of the functions J involved performing a complicated integral over ξ and taking care in handling numerically unstable terms.

F. Application to phantom and real data

In order to demonstrate that the 2D direct-spline inverse of the Radon transform produces images of comparable or superior quality to filtered backprojection (FBP), we reconstructed images of a numerical Hoffman brain phantom [17] using both methods. The sinogram consisted of 128 simulated noiseless projections of the phantom, taken over 360° and each comprising 400 projection bins. The sinogram contained a total of 1.72×10^8 counts. We first reconstructed the phantom using standard area-weighted FBP with a ramp filter (cutoff=1.0 times the Nyquist frequency). We then fit an

interpolating spline to the data at each of the 128 projection angles and used the 2D direct-spline technique to reconstruct.

Poisson noise was then added to the sinogram. Images were reconstructed using FBP with a ramp filter (cutoff=1.0) as well as the direct-spline technique using an interpolating spline in order to see how the algorithms compared in the presence of noise without prior smoothing. Smoothing splines were then fit to the projection data at each angle, using GCV to determine the smoothing parameter, and an image reconstructed from the coefficients using the direct-spline method. In order to examine the performance of FBP in the face of data with the same degree of smoothness, we sampled these smoothing splines to generate a discrete sinogram that was reconstructed using FBP with a ramp filter (cutoff=1.0).

For the 3D case, we reconstructed images of a Data Spectrum ventricular phantom from projection data acquired on a Picker XP3000 three-headed SPECT system fitted with high-resolution, parallel-hole collimators. The phantom was filled with 3.27 mCi of Tc-99m and placed at the center of rotation. Each head collected data on a 128x128 grid and at 120 projection angles over 360°. We rebinned the projection data from a single head to generate planar integrals on a 128x60x120 grid. Images were reconstructed from this planar-integral data using 3D FBP, and also using the 3D direct-spline inversion method after splines were fit to the data as described in section II.E. We then fit smoothing splines to the planar-integral data and reconstructed directly from the spline coefficients. Finally, we sampled the smoothing splines to generate a smoothed, discrete sinogram and used that as input to the 3D FBP algorithm.

G. Resolution, noise, and signal-to-noise studies

G.1. Resolution

In order to compare quantitatively the resolutions of the direct-spline algorithms (both two- and three-dimensional) with their FBP counterparts, we acquired projection images of a small (1 cm) spherical lesion containing 7.6 mCi of Tc-99m placed in an 800 cc cylindrical phantom containing cold (zero-activity) water. A Picker XP2000 two-headed SPECT system fitted with ultra-high-resolution, parallel-hole collimators was used. The heads rotated at their minimum radius of rotation (9 cm) and acquired 120 views over 360° onto a 128x128 matrix (pixel size=4.67 mm).

For the two-dimensional algorithms, we extracted the 2D sinogram corresponding to the slice through the center of the lesion and reconstructed images using FBP with a ramp filter (cutoff=1.0) as well as using the direct spline inversion with interpolating splines. The reconstructed lesion was approximately a symmetric 2D Gaussian in shape and we determined its full-width half-maximum by collapsing it into a one-dimensional function and fitting this profile with a Gaussian curve. For the three-dimensional reconstruction algorithms, we rebinned the projection data to generate planar-integral data on a 128x60x120 grid. We then used 3D FBP and the 3D direct spline method (using interpolating splines) to reconstruct the slice through the center of the lesion and determined the FWHM of the resulting Gaussian by the same method as above.

In order to isolate the contribution of the reconstruction algorithm to the FWHM of the lesion in the reconstructed images, the contribution from the lesion's inherent width as well as the imaging system's point-spread function had to be removed. The net effect of these two factors was estimated by determining the average FWHM of the lesion as it appeared in the 120 projection images. Assuming then that the reconstruction algorithms could be characterized by Gaussian point-spread functions, the FWHM of these functions were determined by subtracting (in quadrature) the average projection FWHM from the FWHMs of the reconstructed lesions discussed above.

G.2. Noise Levels

To characterize the noise level in images reconstructed by the direct-spline methods and their FBP counterparts, we acquired 20 1-minute projection datasets of the same 800 cc cylindrical phantom used in the previous section, this time containing 3.7 mCi of Tc-99m and no lesion. One slice of this uniform cylinder was reconstructed for each of the 20 datasets using 2D FBP, 3D FBP, 2D spline inversion, and 3D spline inversion (both of these using interpolating splines). For a given algorithm, the same six circular regions of interest (ROI) were examined in each of the 20 slices and the coefficient of variation (the standard deviation of the pixel values in the ROI divided by the mean of the pixel values in the ROI) calculated for each of the 120 ROIs. The average of these 120 coefficients of variation was then computed.

G.3. Signal-to-noise ratio

It is not uncommon for a reconstruction algorithm to offer enhanced resolution at the price of amplified noise. The overall effect of such a tradeoff is sometimes better characterized by computing a signal-to-noise ratio (SNR). We used the two datasets described above to compute so-called ideal-observer SNRs. The ideal-observer framework [18] offers a way of assessing the amount of information the data output by an imaging device (possibly modified by image processing) contains with regard to the performance of a specified task. For linear imaging processes in which the noise in the output image is assumed to be additive, Gaussian, zero-mean, stationary, and independent of the presence or absence of the signal, the ideal-observer framework allows us to characterize fully the quality of the imaging system data with respect to the performance of the specified signal-detection task with a single number, the ideal observer SNR. This can be expressed as

$$SNR_I^2 = \int \frac{|\Delta S_{out}(\upsilon)|^2}{W(\upsilon)} d\upsilon, \qquad (17)$$

where $|\Delta S_{out}(\upsilon)|^2$ is the power spectrum of the signal in output space, i.e. after it has been degraded by the imaging system and possibly image processing, and $W(\upsilon)$ is the Wiener spectrum.

In order to calculate the ideal-observer SNR of the reconstructed images, we regard the 20 reconstructed images of the cylinder alone described in section G.2 as an ensemble of so-called background images. By adding the projections of the lesion described in section G.1 with a suitable scaling factor (chosen in this case to produce a 6:1 lesion-background

concentration ratio) to the projections of the cylinder and then reconstructing, we can generate an ensemble of signal-plus-background images. Then $\left|\Delta S_{out}(\upsilon)\right|^2$ is easily determined by computing the power spectrum of the difference between the two ensemble averages. The Wiener spectrum $W(\upsilon)$ can be computed from the ensemble of background images. The detailed procedure for the calculation is as follows:

1. The lesion projections were scaled to simulate a desired lesion-background concentration ratio (6:1 in this case) and added to each of the 20 sets of background projections.

2. Images of the slice through the center of the lesion were reconstructed for the 20 signal-plus-background datasets using eight different methods: the 2D spline-based inversion with interpolating splines, 2D FBP with a ramp filter (cutoff=1.0), the 2D spline-based inversion using smoothing splines, 2D FBP using a sinogram sampled from these smoothing splines, the 3D spline-based inverse using interpolating splines, 3D FBP, the 3D spline-based inverse using smoothing splines, and 3D FBP using a sinogram sampled from these smoothing splines.

3. The 20 corresponding sinograms of background alone were reconstructed in the same eight ways.

4. An average signal image was determined for each reconstruction method by subtracting the average of the 20 background reconstructions from the average of the 20 signal-plus-background reconstructions. The signal power spectrum was computed by squaring the Fourier transform of this image.

5. While SPECT images are not stationary in general, the attenuated projections of a uniform cylinder of this diameter are quite flat over a broad central region, and thus one might reasonably expect the reconstructed images of this cylinder to be locally stationary near their center, precisely where the lesion is expected to lie [19]. A "local" Wiener spectrum in this region was thus computed from the 20 images of background alone by subtracting the average background image from each of the individual background images, resulting in 20 noise images. Each such image was multiplied by a circularly symmetric window of the form:

$$w(\mathbf{r}) = 1 \quad \text{for } |\mathbf{r}| \le 0.9R,$$

$$w(\mathbf{r}) = 0.5 \times (1 + \cos(\pi(|\mathbf{r}| - 0.9R) / 0.2R),$$
$$\text{for } 0.9R < |\mathbf{r}| < 1.1R, \text{ and}$$

$$w(\mathbf{r}) = 0 \quad \text{for } |\mathbf{r}| \ge 1.1R, \tag{19}$$

where R is the radius of the circular region over which the noise is expected to be stationary (chosen to be 6 pixels) and $\mathbf{r}$ the radial position in the image. The power spectrum of each of the 20 images was computed by taking the square of the Fourier transform of the resulting image. The 20 power spectra were then averaged and scaled so that the volume under the Wiener spectrum equaled the average variance in the circle of radius R [20-22].

6. The ideal-observer SNR was then determined by summing the quotient of the calculated signal spectrum and Wiener spectrum.

The ideal-observer SNR of the raw projection data represents an upper bound on the ideal-observer SNR of the reconstructed images, a point that is discussed further in section IV. Because the noise in the projection data is uncorrelated, the ideal-observer SNR can be computed using a simpler expression than equation (17),

$$SNR_i^2 = \Delta s^t \left(diag \left\{ \left\langle p(\xi_i, \varphi_j) \right\rangle \right\} \right)^{-1} \Delta s, \tag{18}$$

where Δs is the signal projection in the spatial domain expressed as a one-dimensional vector, $diag\{\ \}$ denotes a diagonal matrix and $\left\langle p(\xi_i, \varphi_j) \right\rangle$ is the noise-free background projection data [23]. This is estimated by the sample mean of the 20 noisy background projection datasets.

Finally, it should be noted that in using the ideal-observer framework at all it is implicitly being assumed that the data satisfies the assumptions discussed above: that the system is linear and that the noise in the planar or reconstructed images is additive, Gaussian, stationary, zero-mean, and independent of the presence or absence of the signal. Given the reasonably high count levels (~10-15/pixel), the fact that the signal is relatively small and low contrast, and the discussions of stationarity above, these assumption about the noise are not unreasonable. The requirement of linearity seemingly undermines the use of the framework to analyze images that are reconstructed from smoothing splines that have been fit using an adaptive, and thus non-linear algorithm. However, what is truly required for equation (17) to be meaningful is not linearity in the face of any possible input but more specifically that the system transfer function be the same whether the particular signal of interest is present or absent from the particular background of interest. Again, because the signal in question is relatively small and low contrast, it should not greatly affect the noise properties of the projection images and thus the use of smoothing splines should yield a similar effective system transfer function whether the signal is present or absent.
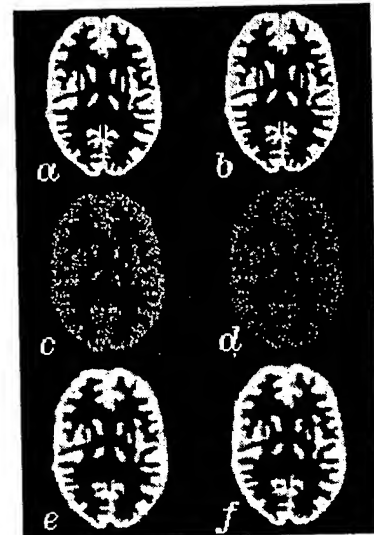

Figure 1. Reconstructions of a Hoffman brain phantom from simulated projections without noise (a-b) and with Poisson noise added (c-f). Reconstruction methods are: (a) FBP with ramp filter (cutoff=1.0), (b) direct-spline inversion using interpolating splines, (c) FBP with ramp filter (cutoff=1.0), (d) direct-spline inversion using interpolating splines, (e) FBP from a sinogram generated by sampling smoothing splines, (f) direct-spline inversion using smoothing splines.

III. RESULTS

The results of reconstructing the Hoffman brain phantom with and without noise using both the 2D direct-spline inverse and 2D FBP are depicted in Figure 1. The algorithms clearly yield qualitatively similar results.

The results of reconstructing the ventricular phantom data using 3D direct-spline inversion with interpolating splines, 3D FBP, 3D direct-spline inversion with smoothing splines, and 3D FBP using a sinogram resampled from the smoothing splines are depicted in Figure 2. The algorithms are again seen to yield qualitatively similar results.

The resolution measurements for the four basic algorithms—2D direct-spline inversion, 2D FBP with a ramp filter (cutoff=1.0), 3D direct-spline inversion, and 3D FBP—are summarized in Table 1. The results indicate that the direct-spline inversions have superior resolution to FBP in both the 2D and 3D cases and also that the 2D algorithms have superior resolution to the 3D algorithms.

The results of the noise study are summarized in Table 2, where it can be seen that the noise level in the direct-spline reconstructions is higher than that in the FBP reconstructions and that the noise level in the 2D reconstructions is higher than that in the 3D reconstructions.

Finally, the ideal-observer signal-to-noise ratio results are summarized in Table 3 for the four basic algorithms as well as their counterparts using smoothing splines. We observe that the direct-spline algorithms have slightly lower SNRs than their FBP counterparts when using interpolating splines, while the SNRs become comparable when using smoothing splines. Furthermore, the use of smoothing splines seems to have little effect on SNR in the 2D case while degrading it in the 3D case. All of the reconstructed images are seen to have lower SNRs than the raw projection data, a fact that is discussed in greater detail in the next section. Typical images reconstructed using each of these eight methods are shown in Figure 3.

Table 1
FWHM of in-plane reconstruction point-spread functions

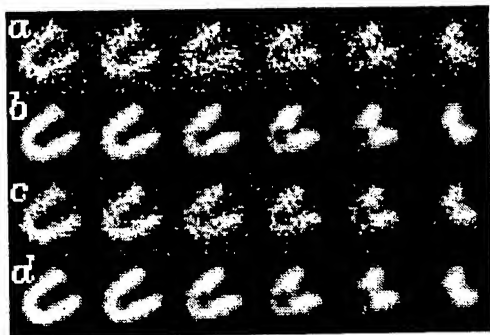| Algorithm	FWHM
2D direct spline	1.6 mm
2D FBP	4.5 mm
3D direct spline	3.9 mm
3D FBP	5.0 mm

Figure 2. Selected slices of a ventricular phantom reconstructed by (a) 3D direct-spline inversion using interpolating splines, (b) 3D direct-spline inversion using smoothing splines, (c) 3D FBP, and (d) 3D FBP from a sinogram obtained by sampling the smoothing splines in (b).

Table 2
Coefficients of variation for various reconstruction algorithms

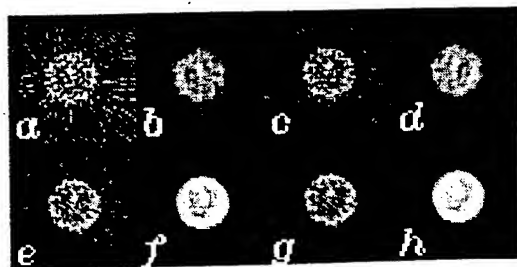| Algorithm	COV
2D direct spline	0.60
2D FBP	0.39
3D direct spline	0.34
3D FBP	0.23

Figure 3. Reconstructions of a selected slice of a cylindrical phantom containing a spherical lesion. Reconstruction methods: (a) 2D direct-spline inversion using interpolating splines, (b) 2D direct-spline inversion using smoothing splines, (c) 2D FBP, (d) 2D FBP from a sinogram obtained by sampling the smoothing splines in (b), (e) 3D direct-spline inversion using interpolating splines, (f) 3D FBP, (g) 3D direct-spline inversion using smoothing splines, (h) 3D FBP from a sinogram obtained by sampling the smoothing splines in (g)

Table 3
Ideal-observer SNRs for various reconstruction algorithms

Algorithm	Ideal-observer SNR
Sinogram	10.7
2D interpolating spline	7.9
2D FBP	8.6
2D smoothing spline	8.4
2D FBP w/ smth. spline	8.4
3D interpolating spline	9.0
3D FBP	9.6
3D smoothing spline	7.5
3D FBP w/ smth. spline	7.0

IV. DISCUSSION AND CONCLUSIONS

As discussed in section I, the principal difference between the 2D and 3D direct-spline inversion algorithms and FBP is

in the nature of the interpolation step upon backprojection. The interpolation in FBP is simply less accurate than the more sophisticated cubic-spline interpolation used in the direct-spline method. The cruder FBP interpolation is more likely to smooth over high-frequency variations in the projection data than is cubic-spline interpolation, and thus it is not surprising that the FBP algorithms have inferior resolution to the spline-based algorithms, as illustrated in Table 1. However, because the high-frequency components of the data include considerable noise as well, the FBP algorithms would be expected to produce less noisy reconstructions than the spline-based reconstructions. This expectation is confirmed by the results of the noise study reported in Table 2.

The 3D direct-spline and FBP algorithms are both seen to have inferior resolution and lower noise levels than their 2D counterparts. This can be attributed to the fact that the 3D reconstruction process involves an additional averaging or smoothing step which occurs when the raw projection data is rebinned into a planar-integral sinogram by performing area-weighted forward projections of the 2D projection data at each projection angle.

Table 3 lists the ideal-observer SNR for the raw projection data and for the various reconstruction approaches. It is a fact that processing or even image reconstruction can never improve the ideal-observer SNR over that found in the raw projection data. However, these operations can certainly *diminish* the SNR if they are in some way singular and if the signal vector has components in the null space. We observe that all the reconstructed image SNRs are in fact lower than that of the raw projection data. The differences among those reconstructed image SNRs give some clue as to how much of the information contained in those projections is preserved in the reconstructed image. For instance, we see that ideal-observer SNRs are slightly lower for the 2D and 3D direct-spline inversions using interpolating splines than for their FBP counterparts using ramp filters. For this particular detection task, then, the interpolating-spline algorithm's amplification of noise outweighs the improvement it affords in resolution relative to FBP. However, when the noise is mitigated prior to reconstruction, as when the projection data has been fitted with smoothing splines, the SNR gap between the spline algorithms and FBP is considerably narrowed in the 2D case and reversed in the 3D case. This suggests that the spline-based algorithms may be of greatest use when resolution is paramount and the data contains relatively little noise, a situation more often encountered in computed tomography than in nuclear medicine.

The use of smoothing splines is seen to provide little or no improvement in SNR in the 2D case. Smoothing does affect the ideal-observer SNRs in the 3D case, but for the worse. It is clear from examining the images of Figure 3 that the reconstructed images using 3D algorithms and smoothing splines have an oversmoothed appearance. This can most likely be attributed to the fact, mentioned above, that the 3D reconstruction process effectively involves a prior smoothing during the rebinning step. The adaptive smoothing algorithm certainly makes allowances for the lower variability in the rebinned data and smooths this data less than it would the raw projection data. However, the modified statistics of the rebinned data simply do not agree as well with the statistical

model assumed by the smoothing algorithm, so it is perhaps not surprising that it yields a sub-optimal result. It remains a topic for further investigation as to whether smoothing the raw projection data prior to the rebinning step produces a better result.

V. ACKNOWLEDGMENTS

VI. REFERENCES

[1] G.N. Hounsfield, "Computerized transverse axial scanning (tomography). I. Description of system," *Br. J. Radiology*, vol. 46, pp. 1023-1047, 1973.

[2] M.M. Ter-Pegossian, M.E. Phelps, E.J. Hoffman and N.A. Mullani, "A positron-emission transaxial tomograph for nuclear imaging (PETT)," *Radiology*, vol. 114, pp. 89-98, 1975.

[3] T.F. Budinger, "Single-photon emission computed tomography," in *Diagnostic Nuclear Medicine*, M.P. Sandler *et al.*, eds., Baltimore: Williams and Wilkens, 1996, pp. 121-138.

[4] H.H. Barrett, "The Radon transform and its applications," in *Progress in Optics XXI*, E. Wolf, ed. Amsterdam: Elsevier Science Publishers, 1984, pp. 217-286.

[5] S. Deans, *The Radon Transform and Some of its Applications*, New York: Wiley, 1983.

[6] G.T. Herman, *Image Reconstruction from Projections: The Fundamentals of Computerized Tomography*, Boston: Academic Press, 1980.

[7] L.A. Shepp and B.F. Logan, "The Fourier reconstruction of a head section," *IEEE Trans. Nucl. Sci.*, vol. 21, pp. 21-43, 1974.

[8] R.H. Huesman, G.T. Gullberg, W.L. Greenberg and T.F. Budinger, *User Manual, Donner Algorithms for Reconstruction Tomography*, Lawrence Berkeley Laboratory, University of California, 1977.

[9] C. Wu, "Fully three-dimensional reconstruction in PET and SPECT by the use of three-dimensional Radon transforms," *PhD Thesis*, University of Chicago, 1994.

[10] G. Wahba, "A new approach to the numerical evaluation of the inverse Radon transform with discrete, noisy data," *Technion Preprint Series* MT-479, 1980.

[11] P.J. Green and B.W. Silverman, *Nonparametric Regression and Generalized Linear Models*, London: Chapman Hall, 1994.

[12] G.T. Herman G T and A. Naparstek, "Fast image reconstruction based on a Radon inversion formula

appropriate for rapidly collected data," *SIAM J. Applied Mathematics*, vol. 33, pp. 511-533, 1977.

[13] J.A. Fessler, "Tomographic reconstruction using information-weighted spline smoothing," in *Information Processing in Medical Imaging*, H.H. Barrett and A.F. Gmitro, eds., Berlin: Springer-Verlag, 1993, pp. 290-300.

[14] D.M. Bates, M.J. Lindstrom, G. Wahba and B. Yandell "GCVPACK routines for generalized cross-validation," *Commun. Statist. Simul. Comput.*, vol. 16, pp. 263-297, 1987.

[15] M.Y. Chiu, H.H. Barrett and R.G. Simpson, "Three-dimensional image reconstruction from projections," *J. Opt. Soc. Am.*, vol. 70, pp. 755-762, 1980.

[16] F. Natterer, *The Mathematics of Computerized Tomography*, Stuttgart: John Wiley & Sons, 1986.

[17] E.J. Hoffman, P.D. Cutler, W.M. Digby and J.C. Mazziotta, "3-D phantom to simulate cerebral blood flow and metabolic images for PET," *IEEE Trans. Nucl. Sci.*, vol. 37, pp. 616-620, 1990.

[18] "Medical imaging—the assessment of image quality," *ICRU Report 54*, 1996. Available from ICRU, 7910 Woodmont Ave., Bethesda, MD, 20814.

[19] S.C. Moore, M.F. Kijewski, S.P. Müller and B.L. Holman, "SPECT image noise power: effects of nonstationary projection noise and attenuation compensation," *JNM*, vol. 29, pp. 1704-1709, 1988.

[20] S.J. Riederer, N.J. Pelc and D.A. Chesler, "The noise power spectrum in computed x-ray tomography," *Phys. Med. Biol.*, vol. 23, pp. 446-454, 1978.

[21] K. Hanson, "Detectability in computed tomographic images," *Med. Phys.*, vol. 6, pp. 441-451, 1979.

[22] M.F. Kijewski and P.F. Judy, "The noise power spectrum of CT Images," *Phys. Med. Biol.*, vol. 32, pp. 565-575, 1987.

[23] J. Yao and H.H. Barrett, "Predicting human performance by a channelized Hotelling observer model," *Proc. SPIE*, vol. 1768, pp. 161-168, 1992.

APPENDIX A

This appendix supplies the details connecting equation (9) to equation (8). Solving the integrals in equation (8) yields

$$J_{r,\theta}(\varphi_j) = \sum_{\substack{i=-N \\ i \neq m}}^{N-1} (b_i + c_i \xi' + d_i \xi'^2) \ln\left(\frac{\xi' - \xi_i}{\xi' - \xi_{i+1}}\right) - \sum_{i=-N}^{N-1} (c_i + 2d_i \xi')(\xi_{i+1} - \xi_i)$$

$$+ \frac{1}{2} \sum_{i=-N}^{N-1} d_i \left[(\xi' - \xi_i)^2 - (\xi' - \xi_{i+1})^2\right] + (b_m + c_m \xi' + d_m \xi'^2) \ln\left(\frac{\xi' - \xi_m}{\xi_{m+1} - \xi'}\right).$$

(A.1)

There are three terms in equation (A.1), containing $\ln[(\xi' - \xi_{m-1})/(\xi' - \xi_m)]$, $\ln[(\xi' - \xi_{m+1})/(\xi' - \xi_{m+2})]$, and $\ln[(\xi' - \xi_m)/(\xi_{m+1} - \xi')]$, respectively, which are numerically unstable when ξ' is near ξ_m in the case of the first term, near ξ_{m+1} in the case of the second, or near either in the case of the third. The three terms, whose sum we denote as T are

$$T = (b_{m-1} + c_{m-1} \xi' + d_{m-1} \xi'^2) \ln\left(\frac{\xi' - \xi_{m-1}}{\xi' - \xi_m}\right) + (b_{m+1} + c_{m+1} \xi' + d_{m+1} \xi'^2) \ln\left(\frac{\xi' - \xi_{m+1}}{\xi' - \xi_{m+2}}\right) + (b_m + c_m \xi' + d_m \xi'^2) \ln\left(\frac{\xi' - \xi_m}{\xi_{m+1} - \xi'}\right).$$

(A.2)

Spline identities that follow from the fact that a cubic spline has continuous first and second derivatives at ξ_m,

$$(b_m - b_{m-1}) + (c_m - c_{m-1})\xi_m + (d_m - d_{m-1})\xi_m^2 = 0$$

(A.3)

$$(c_m - c_{m-1}) + 2(d_m - d_{m-1})\xi_m = 0,$$

and a similar pair reflecting the continuity at ξ_{m+1} can be used to remove the singularities in equation (A.2). Using these identities we can reexpress T as

$$T = (b_{m-1} + c_{m-1} \xi' + d_{m-1} \xi'^2) \ln(\xi' - \xi_{m-1}) + (d_m - d_{m-1})(\xi' - \xi_m)^2 \ln(\xi' - \xi_m)$$

$$+ (d_{m+1} - d_m)(\xi_{m+1} - \xi')^2 \ln(\xi_{m+1} - \xi') - (b_{m+1} + c_{m+1} \xi' + d_{m+1} \xi'^2) \ln(\xi_{m+2} - \xi'),$$

(A.4)

in which the second and third terms pose no numerical instability because $(\xi' - \xi_m)^2 \ln(\xi' - \xi_m) \to 0$ as $\xi' \to \xi_m$ and $(\xi_{m+1} - \xi')^2 \ln(\xi_{m+1} - \xi') \to 0$ as $\xi' \to \xi_{m+1}$.

Hence the final expression for $J_{r,\theta}(\varphi_j)$ is given by

$$J_{r,\theta}(\varphi_j) = \sum_{\substack{i=-N \\ i \neq m-1, m, m+1}}^{N-1} (b_i + c_i \xi' + d_i \xi'^2) \ln\left(\frac{\xi' - \xi_i}{\xi' - \xi_{i+1}}\right) - \sum_{i=-N}^{N-1} (c_i + 2d_i \xi')(\xi_{i+1} - \xi_i) + \frac{1}{2} \sum_{i=-N}^{N-1} d_i \left[(\xi' - \xi_i)^2 - (\xi' - \xi_{i+1})^2\right] + T,$$

(A.5)

where T is given by equation (A.4).

Few-View Tomography Using Interpolating and Smoothing Splines with Implications for Cardiac SPECT

P. J. La Rivière, *Student Member, IEEE,* and X. Pan, *Member, IEEE* Department of Radiology, The University of Chicago, Chicago, IL, 60637

Abstract

The ability to reconstruct high-quality tomographic images from a smaller number of projections than is usually used could reduce imaging time for many nuclear-medicine studies. This would particularly benefit studies such as cardiac SPECT where patient motion during long acquisitions can lead to motion artifacts in the reconstructed images. To this end, we have investigated sinogram pre-processing techniques designed to enable filtered backprojection (FBP) to produce high-quality reconstructions from a small number of views. Each projection is first smoothed by performing roughness-penalized nonparametric regression using a generalized linear model that explicitly accounts for the Poisson nature of the data. The resulting fit curves are natural cubic splines. After smoothing, additional angular views are generated using periodic spline interpolation, and images are reconstructed using FBP. The algorithm was tested on data from SPECT studies of a cardiac phantom placed at various radial offsets to enable examination of the algorithm's dependence on the radial extent of the object being imaged.

I. INTRODUCTION

In routine nuclear-medicine tomographic studies, there is usually a tradeoff between image quality and imaging time. Increasing the number of angular views acquired, the number of counts per view, or both will generally improve image quality but will also lengthen imaging time. In general, concerns about image quality take precedence over concerns about imaging time. However, in studies where the patient must assume an awkward or uncomfortable position, long acquisition times can potentially lead to patient motion and thus to motion artifacts in the reconstructed images. One way to reduce imaging time without significant sacrifice of image quality is to use a continuous acquisition mode, in which the time wasted in moving the camera between views is eliminated. However, continuous acquisition is not recommended for some of the studies most plagued by motion artifacts, such as gated cardiac SPECT. In such studies, reducing imaging time must be accomplished either by reducing the number of angular views or by reducing the number of counts per view. In this paper, we focus on the first approach, investigating algorithms tailored to generate diagnostically useful images from a smaller number of angular views than is usually used while holding the number of counts per view constant.

The minimum number of angular views required to produce an accurate tomographic reconstruction of a given object is dictated by two factors. First, the angular sampling of the object's sinogram must satisfy the Nyquist sampling condition. Second, the number of angular samples must satisfy the reconstruction algorithm's implicit assumptions about the density of angular sampling. For instance, it is well known that images reconstructed from a small number of angular views by filtered backprojection (FBP) are degraded by star-shaped artifacts. These two conditions are not in general equivalent, as can be appreciated most keenly when considering the case of imaging a circularly symmetric object. In this instance, a single projection view is sufficient to satisfy the Nyquist sampling condition, while an FBP reconstruction from this single view would be an uninterpretable set of parallel streaks.

The Nyquist condition is the more fundamental of the two sampling requirements discussed above, because when it is satisfied by a number of samples less than the number required by the reconstruction algorithm, it is in principle possible to interpolate the additional views needed. This is only strictly true in the absence of noise, of course, a condition that rarely obtains in emission tomography. In the presence of noise, angular interpolation usually leads to circular artifacts, because the noise process is rarely if ever bandlimited to the Nyquist frequency of the samples. However, by smoothing each projection prior to interpolation, the noise-engendered interpolation artifacts can be effectively eliminated. We opted to explore a statistical smoothing approach: roughness-penalized non-parametric regression using an explicit Poisson statistical model [1]. This leads to fit functions that are natural cubic splines, piecewise cubic polynomials that are continuous up to and including the second derivative and satisfying the so-called natural boundary conditions. While this approach is similar in spirit to the information-weighted splines of Fessler [2], the use of the explicit Poisson model in the present case leads to a rather different, iterative algorithm. Moreover, rather than choosing the smoothing parameter *a priori,* we have implemented an automatic algorithm for determining it, based on the principle of cross-validation and adapted for Poisson-distributed data. The interpolation of additional views is also performed using splines, this time satisfying periodic boundary conditions. Reconstruction then proceeded as usual, with recognition of the fact that the sinogram has already been smoothed.

II. METHODS

A. Interpolation of angular views

Consider a tomographic acquisition that yields a 2D discrete sinogram $p(\xi_n, \phi_m)$, with $n = 1, \dots, N$ and $m = 1, \dots, M$, and where the number M of angular samples is assumed to satisfy the Nyquist condition at least approximately. We wish to increase the number of angular samples to KM, where K is an integer, by interpolating additional views between the

measured ones. To do this, the sinogram is viewed as a set of 1D sampled functions of projection angle, each labeled by a projection bin ξ_n, with the samples denoted by $p_{\xi_n}(\phi_m)$. Then continuous 1D interpolating curves $p^i_{\xi_n}(\phi)$, where the superscript i indicates interpolated, are fit to each of these sampled functions and resampled to obtain the additional angular views.

Ideally, a periodic interpolation method should be chosen in order to make use of the periodicity of the angular samples. We chose periodic cubic splines, which are accurate interpolating curves comprised of generally different third-order polynomials between each pair of known abscissas ϕ_m and ϕ_{m+1}, with the overall curve being continuous up to and including the second derivative at each abscissa [1]. Naturally, in the case of an interpolating spline, the curve $p^i_{\xi_n}(\phi)$ is also constrained to pass through the known ordinate values $p_{\xi_n}(\phi_m)$ at each abscissa ϕ_m. The spline can be represented as

$$ p^i_{\xi_n}(\phi) = a_m + b_m\phi + c_m\phi^3/2 + d_m\phi^3/3, \qquad (1) $$

for $\phi \in [\phi_m, \phi_{m+1}]$, where $m = 1, \ldots, M$. The process of fitting an interpolating spline is then tantamount to solving a set of linear equations for the defining coefficients a_m, b_m, c_m, and d_m in each interval $[\phi_m, \phi_{m+1}]$ such that the continuity and interpolation conditions are satisfied. For a periodic spline, the coefficients must also satisfy periodic boundary conditions.

B. Smoothing

Prior smoothing of the sinogram is necessary if spline interpolation of angular views is to succeed in the presence of noise. Rather than simply apply a shift-invariant filter to the noisy projection data, we decided to explore a more principled statistical approach using roughness-penalized non-parametric regression based on a generalized linear model (GLM) [1] that explicitly accounts for the Poisson nature of the SPECT data.

1) Non-parametric regression using a GLM

Regression analysis with a single explanatory variable attempts to a curve to a set of data pairs (Y_i, x_i), $(i = 1, \ldots, N)$, where the Y_i are the measured values of the quantity of interest and the x_i the corresponding values of the explanatory variable. The variation in the Y_i is assumed to have two components: a systematic component captured by a vector of predictors θ_i that depends on the x_i, and a random component specifying the distribution of the Y_i given θ_i. In classical linear regression, for example, the systematic component is assumed to be of the form $\theta_i = ax_i + b$, and the Y_i are assumed to be normally distributed about the θ_i.

Nonparametric regression using a GLM relaxes both of the assumptions of classical linear regression. First, it eliminates the assumption that the predictors θ_i depend on the explanatory variable in a simple parametric way, representing them instead as an arbitrary function of the explanatory variables: $\theta_i = g(x_i)$. This would seem to complicate the problem immensely, turning a relatively straightforward finite dimensional estimation problem into an intractable infinite dimensional one. However, in practice the problem is made

tractable by adding the further constraint that the estimated curve $g(x)$ be smooth and by enforcing this constraint by penalizing the likelihood with a term of the form $\int g''(x)^2 dx$. If the unpenalized likelihood depends on $g(x)$ only through its values $g(x_i)$ at the measured points x_i, $(i = 1, \ldots, N)$ (which is usually the case), it can be shown that the minimizer of the penalized likelihood is necessarily a natural cubic spline [1]. These are simply cubic splines that are constrained to be linear outside the set of measured points.

Nonparametric regression using a GLM also relaxes the assumption that the data are normally distributed in favor of the much broader class of exponential distributions, which have probability densities of the form

$$ p(y_i \mid \theta_i, \phi) = \exp\left(\frac{y_i\theta_i - b(\theta_i)}{\phi} + c(y_i, \phi) \right), \qquad (2) $$

where θ_i is the so-called natural parameter of the exponential family and ϕ is a scale parameter [3]. Of particular interest to emission tomography is the choice $b(\theta_i) = e^{\theta_i}$, $\phi = 1$, $c(y_i, \phi) = -\log(y_i!)$, which corresponds to a Poisson distribution with parameter $\lambda_i = e^{\theta_i}$. If a non-parametric dependence $\theta_i = g(x_i)$ of the predictor on the explanatory variables is assumed, the log-likelihood of N independent observations Y_i drawn from the Poisson density is given by

$$ \ell(g, \phi) = \sum_{i=1}^{N} \left(Y_i g(x_i) - \exp\left[g(x_i)\right] - \log(Y_i!) \right). \qquad (3) $$

The goal of roughness-penalized non-parametric regression is to estimate the curve $g(x)$ that maximizes this log-likelihood subject to the penalty $\int g''(x)^2 dx$, i.e. to maximize

$$ \sum_{i=1}^{N} \left\{ Y_i g(x_i) - \exp\left[g(x_i)\right] \right\} - \frac{1}{2}\alpha \int g''(x)^2 dx, \qquad (4) $$

where terms independent of g have been dropped. As mentioned above, this expression can be shown to be maximized by a natural cubic spline and it can also be shown that for natural cubic splines, the penalty $-\frac{1}{2} \int g''(x)^2 dx = -\frac{1}{2}\alpha g^T K g$, where g is an N-element vector with $g_i = g(x_i)$, and K is an NxN matrix with bandwidth 5 that is determined by the spacing of the measurement points x_i. Because the natural cubic spline interpolating any specified set of points $g(x_i)$ is unique, finding $g(x)$ is thus tantamount to finding g [1].

We use Fisher scoring [3] to find the g that maximizes Eq. 4, which yields the following iterative equation

$$ g^{(k+1)} = (W + \alpha K)^{-1} W z^{(k)}, \qquad (5) $$

where, for Poisson data, z is an N-element vector with components $z_i^{(k)} = g_i^{(k)} + \left[Y_i - \exp\left(g_i^{(k)} \right) \right] / \exp\left(g_i^{(k)} \right)$ and W is a diagonal matrix with entries $W_{ii} = \exp\left(g_i^{(k)} \right)$, and the superscript (k) refers to the kth iteration. The initial estimate $g^{(0)}$ is chosen to have components $g_i^{(0)} = \log\{\max(Y_i, \epsilon)\}$,

where ϵ is a small positive introduced to avoid computing log(0). Iteration continues until the sum of the absolute changes in the components of g from one iteration to the next falls below a prespecified threshold. While this would seem to a very computationally intensive procedure, the banded structure of the matrix $(W + \alpha K)^{-1}$ can be exploited to keep the algorithm to $\vartheta(n)$.

2) Choice of the smoothing parameter

The choice of the smoothing parameter α profoundly influences the appearance of the fit curve $g(x)$, for α determines the relative influence of the two terms in the penalized likelihood expression, the first rewarding goodness-of-fit to the data, the second rewarding smoothness. While a reasonable value of α can be found through trial and error for most datasets, a more principled and automatic approach would clearly be preferred.

One automatic approach to choosing the smoothing parameter is based on the principle of cross validation (CV). The approach assumes that the choice of α should yield a fit curve $g(x)$ that accurately predicts the outcomes of further observations. Because one does not necessarily have access to new measurements made under the same conditions as the available ones, CV effectively generates new measurements by omitting each of the actual measurements in turn, fitting a curve to the remaining data by minimizing the penalized log-likelihood for a specified value of α, and comparing the curve's prediction of the omitted measurement with the actual value. One way to express the CV score is [4]

$$CV(\alpha) = \frac{1}{N} \sum_{i=1}^{N} \left\{ \frac{Y_i - g(x_i)}{1 - A_{ii}(\alpha)} \right\}^2, \qquad (6)$$

where A_{ii} are the diagonals of the so-called hat matrix A, which links the values of the estimate at the measured points to the values of the observations at those points: $\mathbf{g} = A(\alpha)\mathbf{Y}$. The α minimizing Eq. 6 can generally be found fairly quickly using a golden section search minimization approach.

The cross-validation score of Eq. 6, based on a residual sum of squares, is more appropriate for normally distributed data than for the more general class of distributions encompassed by GLMs. For GLMs, we follow O'Sullivan *et al.* [5], who propose replacing the residual sum of squares with the generalized Pearson χ^2 statistic such that

$$CV_{\mathrm{GLM}}(\alpha) = \frac{1}{N} \sum_{i=1}^{N} \frac{(Y_i - \widehat{\mu}_i)^2/\widehat{V}_i}{(1 - A_{ii}(\alpha))^2}, \qquad (7)$$

where $\widehat{\mu}_i$ is the estimated mean at x_i, $\widehat{V}_i$ the estimated variance, and both of these, as well as the A_{ii} are computed at the final iteration. The matrix A is given by $(W + \alpha K)^{-1}W$. For Poisson-distributed data, $\widehat{\mu}_i = \widehat{V}_i = \exp(g_i)$.

Calculating the CV score would seem to be an $\vartheta(n^2)$ operation, for computing the matrix A involves inverting the matrix $(W + \alpha K)$. However, Hutchinson and de Hoog have developed an algorithm that allows the diagonal elements of

the inverse of a band matrix to be computed in $\vartheta(n)$, and this is the approach we have used [6].

C. Data Acquisition and Processing

In order to test these algorithms, we acquired projections of a Data Spectrum ventricular phantom filled with 121 MBq (3.27 mCi) of Tc-99m and containing a 1-cm defect insert. The phantom was not placed within a water-filled torso phantom. We imaged this phantom with a Picker 3000XP three-headed SPECT system fit with low-energy, high-resolution, parallel-hole collimators. We used a ^{57}Co source to locate the center of rotation (COR) and then acquired SPECT studies containing 120 angular views over 360° of the phantom placed at five different radial offsets from the COR: 0, 5, 9, 12, and 15 cm. This was done to allow examination of sensitivity of the interpolation technique to the radial extent of the object being imaged. We used a 25-cm radius circular orbit and step-and-shoot mode for all of the acquisitions; each head acquired to a 128x128 pixel image, though we preserved only the 32 slices spanning the phantom. A total of about 500,000 counts was collected. From this data we extracted 3D sinograms corresponding to 15, 30, 60, and 120 views, respectively. Thus we had 20 different sinograms, corresponding to the 20 combinations of radial offset and number of angular views. We reconstructed images from these 20 sinograms using four different processing techniques:

1. No pre-smoothing of the sinogram and slice-by-slice reconstruction from available views by FBP using a Hanning filter (cutoff=0.4).

2. No pre-smoothing of the sinogram, spline interpolation from the available views to 120 angular views, and slice-by-slice reconstruction by FBP using a Hanning filter (cutoff=0.4).

3. Roughness-penalized non-parametric regression smoothing of the sinogram and slice-by-slice reconstruction from the available views by FBP using a ramp filter (cutoff=0.5).

4. Roughness-penalized non-parametric regression smoothing of the sinogram, spline interpolation from the available views to 120 views, and slice-by-slice reconstruction by FBP using a ramp filter (cutoff=0.5).

The particular application of the non-parametric regression technique to this data requires some explanation. The projection data for each of the radial offsets is a 3D sinogram of 128 bins, 32 slices, and 120 angles. While the smoothing technique was applied to each projection in each slice independently, we decided it would be wise to use the same smoothing parameter for all the projections. To do otherwise would have invited inconsistencies in the smoothed sinogram likely to produce artifacts in the reconstructed images. At the same, we wished to select the smoothing parameter to be applied to the data using some form of cross-validation. The solution was to "string" together all of the 1D projections in

the complete 3D sinogram for each radial offset into a single, long 1D function, and to find the smoothing parameter that minimized the CV score for that function when smoothed using the roughness-penalized non-parametric technique. This value of α was then used in smoothing each of the projections in this sinogram individually.

III. Results

For ease of comparison, we have grouped the reconstructed images by the number of angles in the original sinogram, and we show in Fig. 1 the results for three of the radial offsets: 0, 9, and 15 cm. For each combination of number of angles and radial offset we show the results of reconstructing using the four techniques outlined above.

We observe that reconstructions from available views without pre-smoothing or interpolation display star-shaped artifacts and a mottled appearance when the number of views is small. Interpolation alone mitigates the star-shaped artifacts but leads to severe circular artifacts, particularly in the case of a small number of views and a large radial offset. Smoothing alone reduces the noise visibility but has little effect on the star-shaped artifacts. The combination of smoothing and interpolation still produces circular interpolation artifacts in the case of a large radial offset combined with a small number of original views, but these are less severe than when interpolation alone was used. Overall, though, visually appealing reconstructions result for less challenging combinations of radial offset and number of views, including as few as 15 angles in the 0 cm offset case.

While Fig. 1 demonstrates that the algorithms can produce visually satisfactory reconstructions of the cardiac phantom from relatively small numbers of views, the critical question is whether this can be achieved without hindering the detection of small perfusion defects. To answer this we generated bullseye plots from each set of reconstructions. The bullseye plots corresponding to reconstructions from few-view sinograms that have been processed by the spline smoothing and interpolation techniques are shown in Fig. 2, along with bullseye plots corresponding to 120-angle sinograms reconstructed without spline processing. Our phantom contained a 1-cm perfusion defect insert, which produces a depression that is well resolved in many of the plots.

It is clear from these bullseye plots that the defect remains detectable for as few as 15 angles in the case of the 0 cm offset, as few as 30 angles in the case of the 9 cm offset, and as few as 60 angles in the case of the 15 cm offset. These findings correlate well with the visual appearances of the images in Fig. 1. The plots in which the defect is not visible correspond to the images in which severe interpolation artifacts are evident, and thus to the situation case when the number of angular samples strongly fails to meet the Nyquist condition.

IV. Discussion and Conclusions

We have presented a sinogram pre-processing technique combining spline-based smoothing and interpolation

that enables FBP to produce high-quality tomographic reconstructions from a smaller number of views than usual. The technique is applicable to situations where the number of views needed to satisfy the Nyquist condition on the sampling of the angular part of the sinogram is less than the number of views required by FBP to produce artifact-free images. In this situation, we first smooth each 1D projection using a roughness-penalized non-parametric regression approach based on a GLM that explicitly models the Poisson nature of the measured data. One-dimensional periodic interpolation splines are then fit in the angular direction and resampled to generate additional views. Ramp-filtered FBP is than applied to the interpolated sinogram.

Because it is not possible to derive closed-form expressions for the Nyquist and FBP sampling requirements of non-circularly symmetric objects, we have tested the limits of the algorithm by conducting an experimental investigation performing reconstructions from various numbers of projections of a cardiac phantom placed at various radial offsets from the center of rotation of a SPECT system. As expected, the ability of the algorithm to produce visually appealing images that still captured small perfusion defects breaks down for large radial offsets combined with small numbers of starting views. This corresponds to the situation when the number of initial angular samples strongly fails to satisfy the Nyquist condition, a situation that no amount of data processing can remedy.

V. Acknowledgment

VI. References

[1] P. J. Green and B. W. Silverman, *Nonparametric Regression and Generalized Linear Models*. London: Chapman Hall, 1994.

[2] J. A. Fessler, "Tomographic reconstruction using information-weighted spline smoothing," in *Information Processing in Medical Imaging* (H. H. Barrett and A. F. Gmitro, eds.), pp. 290–300, Berlin: Springer-Verlag, 1993.

[3] P. McCullagh and J. A. Nelder, *Generalized Linear Models*. London: Chapman Hall, 1983.

[4] G. Wahba, *Spline Models for Observational Data*. Philadelphia: SIAM Press, 1990.

[5] F. O'Sullivan, B. S. Yandell, and W. J. Raynor, Jr., "Automatic smoothing of regression functions in generalized linear models," *J. Amer. Stat. Assoc.*, vol. 81, pp. 96–103, 1986.

[6] M. F. Hutchinson and F. R. de Hoog, "Smoothing noisy data with spline functions," *Numer. Math.*, vol. 47, pp. 99–106, 1985.
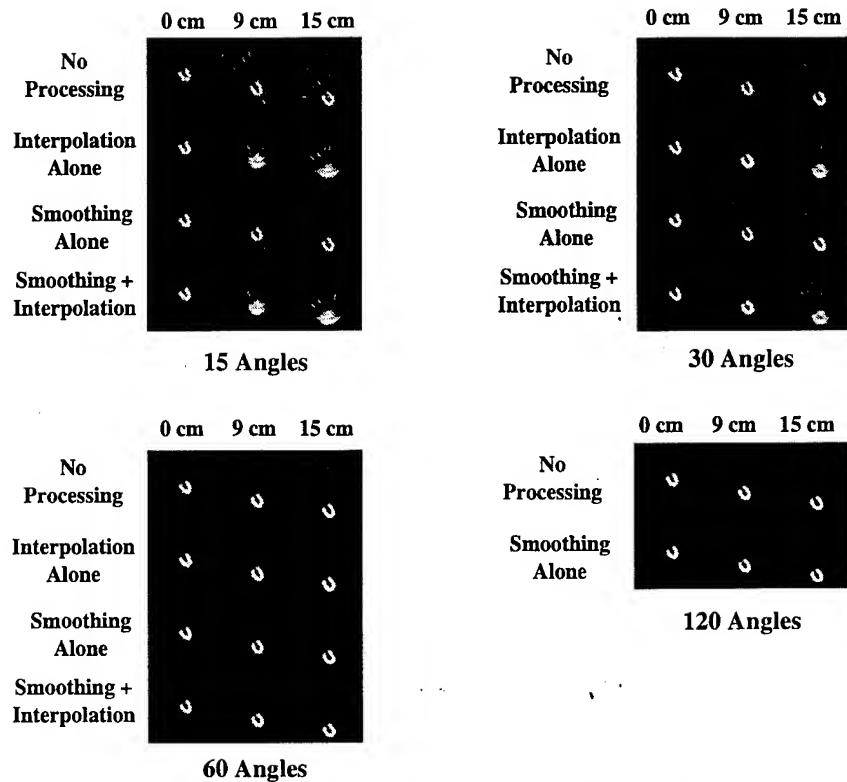
Figure 1: Reconstructions of a representative slice of a cardiac phantom image at three different radial offsets using four different pre-processing approaches and 15, 30, 60, and 120 projection angles.
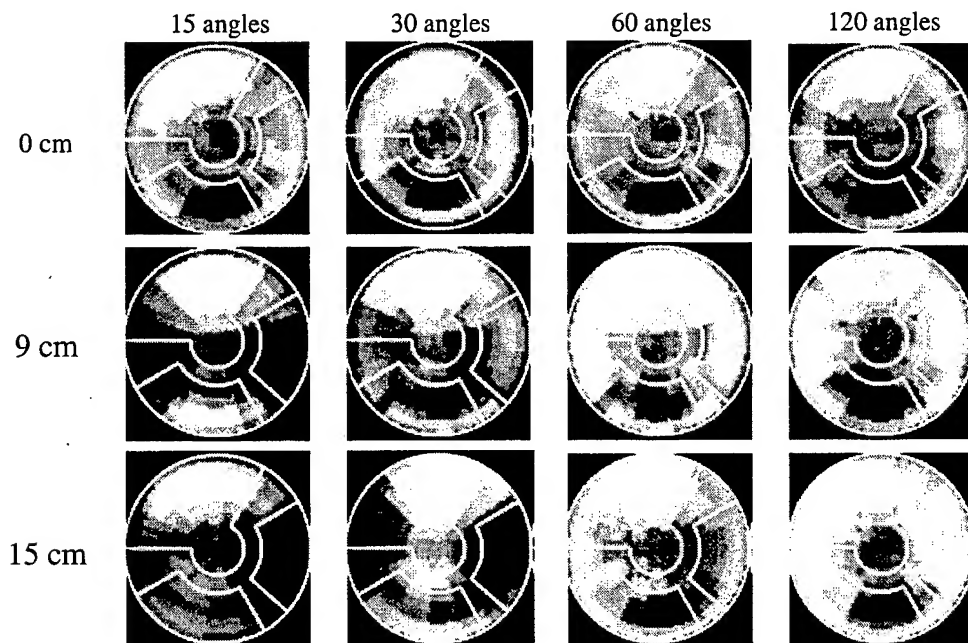

Figure 2: Bulleye plots for the cardiac phantom placed at various radial offsets and for four different numbers of projection angles. The reconstructions from 15, 30, and 60 angles used the non-parametric regression smoothing technique followed by periodic spline interpolation to 120 views prior to reconstruction by ramp-filtered FBP. The reconstructions from 120 angles simply entailed Hanning-filtered FBP.

Noise Properties of Periodic Interpolation Methods with Implications for Few-View Tomography

P. J. La Rivière, *Student Member, IEEE,* and X. Pan, *Member, IEEE* Department of Radiology, The University of Chicago, Chicago, IL, 60637

Abstract

A number of methods exist specifically for the interpolation of periodic functions from a finite number of samples. When the samples are known exactly, exact interpolation is possible under certain conditions, such as when the function is bandlimited to the Nyquist frequency of the samples. However, when the samples are corrupted by noise, it is just as important to consider the noise properties of the resulting interpolated curve as it is to consider its accuracy. In this work, we derive analytic expressions for the covariance and variance of curves interpolated by three periodic interpolation methods—circular sampling theorem, zero-padding, and periodic spline interpolation—when the samples are corrupted by additive, zero-mean noise. We perform empirical studies for the special cases of white and Poisson noise and find the results to be in agreement with the analytic derivations. The implications of these findings for few-view tomography are also discussed.

I. INTRODUCTION

The need to interpolate samples of periodic functions arises in a number of important medical imaging applications. For instance, when performing emission computed tomography imaging of a compact, reasonably symmetric object, such as the breast, one can achieve adequate angular sampling of the object's sinogram with a relatively small number of projection views. However, using filtered backprojection (FBP) to reconstruct the image may still lead to star-shaped artifacts, because FBP implicitly requires a relatively high density of angular samples [1]. In these situations, periodic interpolation may be used to interpolate additional angular views between the measured ones in order to satisfy FBP's sampling requirements. The need to perform periodic interpolation also arises in direct Fourier image reconstruction, where it is necessary to interpolate from a polar to a Cartesian grid in Fourier space [2]. The interpolation is often accomplished using separate 1D interpolations in the radial and azimuthal directions, and the azimuthal interpolation should rightly be periodic.

There exist a number of methods for interpolating periodic functions. Circular sampling theorem (CST) interpolation, for one, is a special case of Whittaker-Shannon (W-S) sinc interpolation that applies to periodic functions [3, 4]. Consider a periodic function $g(x)$ that has period X and which is bandlimited to frequency K (i.e., the coefficients of expansion a_k of the function's Fourier series satisfy $a_k = 0$ for $|k| > K$). Given $N \geq 2K + 1$ samples of $g(x)$ taken at points $x_n = nX/N$ ($n = 0, \ldots, N - 1$) evenly spaced over one period, the CST states [4] that $g(x)$ can be interpolated exactly

by use of

$$g(x) = \sum_{n=0}^{N-1} g(x_n)\sigma_N(x - x_n), \tag{1}$$

where

$$\sigma_N(x) = \sin[(2K + 1)\pi x/X]/N\sin(\pi x/X). \tag{2}$$

If the Nyquist condition is not satisfied, that is, if $g(x)$ is not truly bandlimited to frequency K or if $N < 2K + 1$, Eqs. 1 and 2 no longer represent exact interpolation, but they remain mathematically meaningful and, often, practically useful. For instance, if the spectral components beyond frequency K are negligibly small but not exactly zero, interpolating with Eqs. 1 and 2 remains very accurate.

A second periodic interpolation approach, zero-padding (ZP) interpolation, involves extending the discrete Fourier transform (DFT) of a finite sequence with zeroes and taking an inverse DFT to generate a more densely sampled version of the original sequence with values interpolated at intermediate positions between the original measured samples [5–8]. Specifically, one begins by taking the DFT of the sequence $g(x_n)$, which is given by

$$c_k = \frac{1}{N} \sum_{n=0}^{N-1} g(x_n) \exp(-j2\pi nk/N), \tag{3}$$

for $k = 0, \ldots, N - 1$, where $j = \sqrt{-1}$. Zero-padding involves the creation of a new sequence $d_{k'}$, having $L = PN$ elements (where P is an integer). If $g(x)$ is assumed to be bandlimited to frequency K and if $N \geq 2K + 1$, the sequence $d_{k'}$ is defined as follows:

$$d_{k'} = \begin{cases} c_{k'} & k' = 0, \ldots, K \\ 0 & k' = K + 1, \ldots, L - K - 1 \\ c_{k'-L+N} & k' = L - K, \ldots, L - 1. \end{cases} \tag{4}$$

A more densely sampled sequence $g(x_l)$, where $x_l = lX/L$ ($l = 0, \ldots, L - 1$), is now obtained by taking the inverse DFT of the sequence $d_{k'}$,

$$g(x_l) = \sum_{k'=0}^{L-1} d_{k'} \exp(j2\pi lk'/L), \tag{5}$$

for $l = 0, \ldots, L - 1$. ZP interpolation is generally viewed as a somewhat crude interpolation approach, but in fact nothing could be further from the truth. It can be shown that ZP interpolation is equivalent to CST interpolation, in that the spatial-domain interpolation function corresponding to the ZP

operation in frequency space is just $\sigma_N(x)$ given in (2) [9]. That is, it can be shown that

$$g(x_l) = \sum_{n=0}^{N-1} g(x_n)\sigma_N(x_l - x_n), \qquad (6)$$

where $\sigma_N(x)$ is given by Eq. 2. This equivalence holds regardless of whether the Nyquist condition is satisfied. If it is satisfied, then ZP interpolation, like CST interpolation, is exact. Obviously, the CST interpolation formula can be used to estimate $g(x)$ at any arbitrary point x whereas zero-padding interpolation is constrained to interpolate onto a fixed, equispaced grid P times denser than the original samples, but since in the limit as $P \to \infty$, ZP yields a continuous interpolated curve, we shall treat these two approaches as one in the subsequent analysis, using the continuous form of Eq. 1 to represent both of them.

The final periodic interpolation method to be examined is periodic spline (PS) interpolation. Splines are piecewise cubic polynomials that are continuous up to and including the second derivative at the joints between pieces [10, 11]. Periodic splines are further constrained to satisfy periodic boundary conditions. A spline $\widehat{g}(x)$ can be represented by

$$\widehat{g}(x) = a_n + b_n(x - x_n) + c_n(x - x_n)^2 + d_n(x - x_n)^3, \qquad (7)$$

for $x \in [x_n, x_{n+1}]$, where the x_n are the abscissas at which the data is measured and $n = 0, \ldots, N - 1$. Fitting a spline is thus tantamount to finding the coefficients a_n, b_n, c_n, and d_n subject to interpolation, continuity, and boundary conditions. Because evaluating a spline at a particular point x involves evaluating Eq. 7 using the a_n, b_n, c_n, and d_n corresponding to the interval $[x_n, x_{n+1}]$ in which x falls, we can think of representing a spline as an N-component vector of functions, in which the nth component corresponds to the interval $[x_n, x_{n+1}]$ and contains a function of the form of Eq. 7 with the appropriate values substituted for a_n, b_n, c_n, and d_n. To reflect this understanding and simplify later calculations we introduce the somewhat unconventional notation

$$\widehat{\mathbf{g}}(x) = \mathbf{a} + \mathcal{D}(x)\mathbf{b} + \mathcal{D}^2(x)\mathbf{c} + \mathcal{D}^3(x)\mathbf{d}, \qquad (8)$$

where $\widehat{\mathbf{g}}(x)$ is an N-element vector of functions, $\mathbf{a}$ is the N-element vector with coefficients a_n and likewise for $\mathbf{b}$, $\mathbf{c}$, and $\mathbf{d}$, and $\mathcal{D}$ is a diagonal matrix with $\mathcal{D}_{nn} = x - x_n$.

In fitting a spline, the coefficients are obtained by linear operations on the measured data. If the measured samples $g(x_n)$ are represented as an N-element vector $\mathbf{g}$, the vectors of coefficients can be found from $\mathbf{g}$ through matrix multiplications $\mathbf{a} = A\mathbf{g}$, $\mathbf{b} = B\mathbf{g}$, $\mathbf{c} = C\mathbf{g}$, and $\mathbf{d} = D\mathbf{g}$, where the matrices A, B, C, and D can be deduced from [11]. Substituting for $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, and $\mathbf{d}$ in (8) in terms of these matrix products yields

$$\widehat{\mathbf{g}}(x) = \left[A + \mathcal{D}(x)B + \mathcal{D}^2(x)C + \mathcal{D}^3(x)D\right]\mathbf{g}. \qquad (9)$$

While questions about the accuracy of various interpolation approaches are paramount when the measured samples are known exactly, other concerns arise when the samples are known to be corrupted by noise. In particular, it becomes important to analyze how the noise is propagated into the interpolated samples. Consider now that the samples of $g(x)$ are corrupted by additive, zero-mean noise. These noisy samples can be represented as

$$g(x_n) = \langle g(x_n) \rangle + n(x_n), \qquad (10)$$

where $g(x_n)$ and $n(x_n)$ are random variables, the latter representing zero-mean additive noise, and $\langle \rangle$ represents the expectation operator. The aim of this paper is to derive the covariance and variance of the curves interpolated by means of CST/ZP interpolation and PS interpolation, to compare these analytic predictions with results of Monte Carlo simulations, and to draw conclusions from these results about the suitability of the various approaches for the interpolation task encountered in few-view tomography.

II. METHODS

A. Analytic Derivations

Let $\widehat{g}(x)$ be a curve interpolated from the noisy samples of Eq. 10 by CST or ZP interpolation. The covariance between two points x and x' of this function is given by

$$\text{cov}(x, x') = \langle [\widehat{g}(x) - \langle \widehat{g}(x) \rangle][\widehat{g}(x') - \langle \widehat{g}(x') \rangle] \rangle. \qquad (11)$$

Of course, once the covariance has been computed, the variance at any point x is given by $\text{var}(x) = \text{cov}(x, x)$. Evaluation of Eq. 11 is straightforward as $\widehat{g}(x)$ is given by

$$\widehat{g}(x) = \sum_{n=0}^{N-1} (\langle g(x_n) \rangle + n(x_n))\sigma_N(x - x_n), \qquad (12)$$

where $\sigma_N(x)$ is given by Eq. 2, and thus

$$\text{cov}(x, x') = \left\langle \left[\sum_{n=0}^{N-1} n(x_n)\sigma_N(x - x_n)\right] \right.$$
$$\left. \left[\sum_{m=0}^{N-1} n(x_m)\sigma_N(x' - x_m)\right] \right\rangle, \qquad (13)$$

which can be rewritten

$$\text{cov}(x, x') = \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} [\sigma_N(x - x_n)\sigma_N(x' - x_m)$$
$$\langle n(x_n)n(x_m) \rangle]. \qquad (14)$$

Without additional assumptions about the noise, this expression cannot be simplified any further. In Section III we will consider simplifications to this expression for white noise and uncorrelated Poisson noise.

For periodic spline interpolation, the calculation of a covariance function is complicated by the fact that the points x and x' at which the covariance is being evaluated will in general fall into different intervals of the spline and the spline's value at these points will thus be specified by functions corresponding to different elements of the vector $\widehat{g}(x)$. This is best handled by thinking in terms of computing an NxN
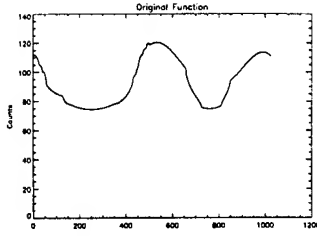
Figure 1: This periodic function, which is the angular function corresponding to one bin of a Shepp-Logan head phantom sinogram, was sampled 128 times and the samples then contaminated by Gaussian or Poisson noise prior to using CST/ZP and periodic spline interpolation to interpolate a sequence 8 times as dense.

matrix of covariance functions whose elements correspond to the possible combinations of pairs of intervals into which x and x' could fall. This matrix, which we denote with capital letters $\text{COV}(x, x')$, is given by

$$\text{COV}(x, x') = \left\langle [\widehat{g}(x) - \langle \widehat{g}(x) \rangle] [\widehat{g}(x') - \langle \widehat{g}(x') \rangle]^T \right\rangle, \quad (15)$$

or using Eq. 9

$$\text{COV}(x, x') = \left\langle \left[A + \mathcal{D}(x)B + \mathcal{D}^2(x)C + \mathcal{D}^3(x)D \right] \mathbf{n} \right.$$
$$\left. \mathbf{n}^T \left[A + \mathcal{D}(x')B + \mathcal{D}^2(x')C + \mathcal{D}^3(x')D \right]^T \right\rangle, \quad (16)$$

where $\mathbf{n}$ is the N-element vector with entries $n(x_n)$, which of course reduces to

$$\text{COV}(x, x') = \left[A + \mathcal{D}(x)B + \mathcal{D}^2(x)C + \mathcal{D}^3(x)D \right]$$
$$\langle \mathbf{n}\mathbf{n}^T \rangle \left[A + \mathcal{D}(x')B + \mathcal{D}^2(x')C + \mathcal{D}^3(x')D \right]^T \quad (17)$$

Again, without further assumptions about the noise, this expression cannot be simplified any further. In Section III we will consider simplifications to this expression for white noise and uncorrelated Poisson noise.

B. Monte Carlo Simulations

To confirm the analytic results derived above and applied below to the cases and white and uncorrelated Poisson, we performed Monte Carlo simulations employing each of these kinds of noise. We first calculated analytically 128 equispaced samples of a typical periodic function encountered in tomography: the angular function corresponding to one bin of the sinogram of a Shepp-Logan head phantom. This function is shown in Fig. 1. We generated 50,000 noise realizations of these samples contaminated with additive, zero-mean Gaussian noise ($\sigma_o^2 = 16$) and 50,000 others with Poisson noise. We then interpolated each of these realizations to 1024 samples using the methods discussed above. For CST and ZP interpolation, we took $K = 63$, the largest value it could have while still being below the Nyquist frequency of the 128 samples. We calculated the sample variance $\widehat{\sigma}^2(x_j)$ at each point x_j and the sample covariance between each of 9 samples and the 1024 interpolated samples.

III. RESULTS

For CST and ZP interpolation, if the noise in the measured samples is assumed to be white, with constant variance σ_o^2 at
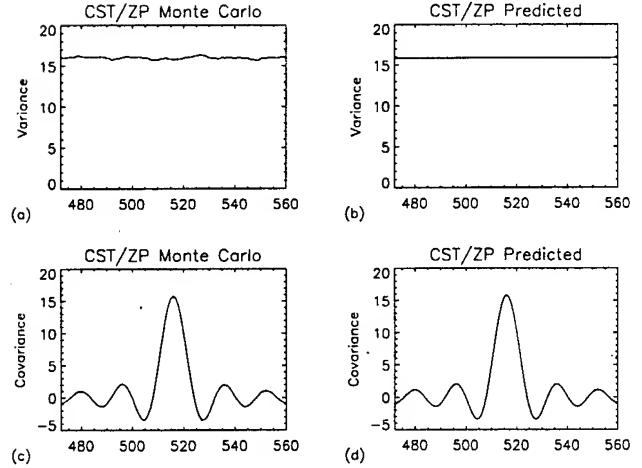

Figure 2: Empirical and predicted variance and covariance using CST/ZP interpolation on 128 samples corrupted by additive, zero-mean Gaussian noise with $\sigma_o^2 = 16$. 1024 samples were interpolated, but to emphasize detail, only a subrange of the resulting variance and covariance curves are shown. The fixed point for the covariance depicted was located midway between two measured points and approximately in the center of the curve as well.

every point, then in Eq. 14, $\langle n(x_n)n(x_m) \rangle = \sigma_o^2 \delta_{nm}$, where δ_{nm} is the Kronecker delta function and we may rewrite Eq. 14 as

$$\text{cov}(x, x') = \sigma_o^2 \sum_{n=0}^{N-1} \sigma_N(x - x_n)\sigma_N(x' - x_n). \quad (18)$$

This expression may be further simplified by comparing the sum with Eq. 1 and realizing that it can be viewed as a CST interpolation of the function $\sigma_N(x - x')$ sampled at points $x' = x_n$. Because $\sigma_N(x - x')$ is periodic and bandlimited to a frequency $K < N/2$, the interpolation is exact, and thus we conclude that

$$\text{cov}(x, x') = \sigma_o^2 \sigma_N(x - x'). \quad (19)$$

From this result, the variance is easily obtained:

$$\text{var}(x) = \text{cov}(x, x) = \sigma_o^2 \sigma_N(0) \quad (20)$$

From the definition of $\sigma_N(x)$ in Eq. 2, we see that $\sigma_N(0) = (2K+1)/N$ and thus conclude that when applying CST and ZP interpolation to samples corrupted by white noise, the variance of the interpolated curve is constant everywhere and equal to $(2K+1)/N$ times the variance in the original samples. Because $N \geq 2K + 1$, this factor is less than or equal to 1, with equality when N minimally satisfies the Nyquist condition.

The Monte Carlo studies support the analytic findings of Eqs. 19 and 20. The upper lefthand plot of Fig. 2 depicts a portion of the 1024-point sample variance curve for CST and ZP interpolation. For comparison, the analytic prediction is shown as well. The sample variance is seen to be approximately constant at all points, in agreement with the analytic prediction, and it is found to have average value 15.933, in agreement with the analytic prediction of $16 * 127/128 = 15.875$. The lower two plots depict the sample and predicted covariance relative to
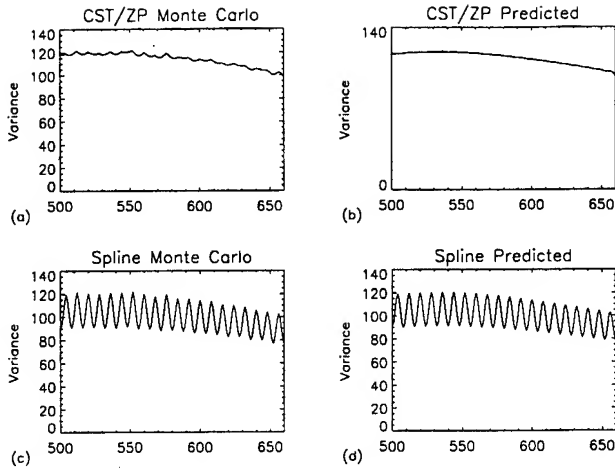
Figure 3: Empirical and predicted variance using CST/ZP interpolation and periodic spline interpolation on 128 samples corrupted by Poisson noise. Again, to emphasize detail, only a subrange of the resulting variance curves is shown.
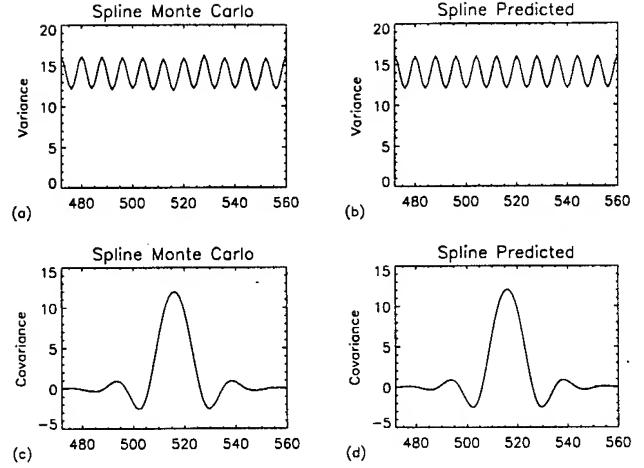

Figure 4: Empirical and predicted variance and covariance using periodic spline interpolation on 128 samples corrupted by additive, zero-mean Gaussian noise with $\sigma_o^2 = 16$. 1024 samples were interpolated, but to emphasize detail, only a subrange of the resulting variance and covariance curves are shown. The fixed point for the covariance depicted was located midway between two measured points and approximately in the center of the curve as well.

an interpolated point located midway between two measured samples. While they are not shown, the sample covariance curves relative to points at other positions in the interval between measured points were found to have essentially identical shapes. This is as expected from examination of the analytic prediction for the covariance function, which depends only on $x - x'$ and is thus shift invariant.

For uncorrelated Poisson noise, $\langle n(x_n)n(x_m)\rangle = \langle g(x_n)\rangle \delta_{nm}$. Thus

$$\text{cov}(x, x') = \sum_{n=0}^{N-1} \langle g(x_n)\rangle \sigma_N(x - x_n)\sigma_N(x' - x_n). \quad (21)$$

This sum can be viewed as a CST interpolation, this time of the function $\langle g(x')\rangle \sigma_N(x - x')$. However, this function, while periodic, is not exactly bandlimited unless $\langle g(x')\rangle$ is constant. However, if $\langle g(x')\rangle$ is slowly varying, at least over the interval between x and x', it can be assumed that the product is approximately bandlimited and thus that

$$\text{cov}(x, x') \cong \langle g(x')\rangle \sigma_N(x - x'). \quad (22)$$

The variance is again easily obtained:

$$\text{var}(x) = \text{cov}(x, x) \cong \langle g(x)\rangle \sigma_N(0). \quad (23)$$

Because $\sigma_N(0) = (2K + 1)/N$, we find that the variance is flat locally (certainly between measured samples), while globally it tracks the variance in the measured samples.

Once again, the Monte Carlo studies confirm the analytic predictions of Eqs. 22 and 23. The top two plots of Fig. 3 illustrate how the variance in the interpolated function, while locally constant, follows the variance in the measured samples over larger distances. The covariance function (not shown) is found to behave over short distances much as it did in the white noise case.

For spline interpolation, the assumption of white noise with variance σ_o^2 implies that in Eq. 17, the outer product

$\langle \mathbf{n}^T \mathbf{n} \rangle = \sigma_o^2 I$, where I is the identity matrix. The properties of the covariance expression given in Eq. 17 cannot be discerned by inspection and the result must be evaluated numerically. In particular, to obtain the variance of the interpolated function, the diagonal elements of the covariance matrix must be evaluated at $x = x'$. The top righthand plot of Fig. 4 illustrates the result of such an evaluation. It is clearly seen that unlike CST and ZP interpolation, the variance is not constant at interpolated points, but in fact falls smoothly to a minimum at points midway between the original measured samples, while remaining unchanged at the measured points. The Monte Carlo results shown in the top lefthand plot of Fig. 4 conform quite closely to this prediction. The two lower plots of Fig. 4 show a portion of the covariance function relative to an interpolated point located midway between two measured samples. It is seen to have a slightly wider central lobe than the equivalent curve for CST/ZP interpolation but also to die out more quickly beyond this central lobe. Unlike in the CST/ZP case, the covariance curves relative to interpolated points in other positions in the intervals between measured points were found to have quite different shapes, becoming asymmetric as the fixed point moves away from the center of an interval.

For Poisson noise and spline interpolation, the same trends are observed. The variance in the interpolated curve behaves locally as it did in the white noise case, falling smoothly to a minimum between measured samples, while globally it tracks the variance in the measured samples. This is clearly seen in the two lower plots of Fig. 3. The covariance curves were found to behave locally as they did in white noise case, changing form as the fixed point is swept through an interval between measured points.

IV. Discussion and Conclusions

The differences between the CST/ZP and periodic spline interpolation approaches emerge most clearly in the stationary white noise case discussed above. Here we saw that CST/ZP interpolation results in a curve with constant variance at all points (a factor of $(2K + 1)/N$ times the variance in the original measured samples), and having a covariance function that depends only on the distance $x - x'$ between two points. Put simply, if the noise in the measured samples is wide-sense stationary, it remains so after interpolation by CST/ZP. (Naturally, the noise does not remain uncorrelated.) This is decidedly not the case for periodic spline interpolation. Here we saw that the variance remains equal to the variance in the original samples only at the interpolated points corresponding to those samples, and that it falls to a minimum at the midpoint between two such samples. Moreover, it was observed that the covariance function does not simply depend on the distance $x - x'$ between two points but depends also on where the positions of the two points within their respective intervals. Thus if the noise in the measured samples is wide-sense stationary, it does not remain so after periodic spline interpolation. Nor, of course, does it remain uncorrelated.

The behavior was similar in the case of Poisson noise, though it cannot be summarized so precisely because Poisson noise is not, in general, wide-sense stationary. In this case, though, CST/ZP interpolation was seen to result in a curve with locally flat variance (i.e., between measured samples), while globally tracking the variance of the measured samples. The covariance also behaved locally as it had in the white noise case—depending only on $x - x'$. For periodic spline interpolation, the variance was again seen to dip locally to minima between measured samples, while globally tracking the variance in the measured samples. The covariance function again exhibited a dependence on the positions of the points x an x' as well as their difference.

The decision to use one approach or the other for the interpolation of additional angular views in few-view tomography should take into consideration the particular requirements of the study being performed. Given that the interpolatory accuracy of the two approaches is generally comparable, at first glance it would seem that the variance-reducing properties of the periodic spline interpolation might be preferable. However, this comes at the cost of having widely varying variance levels in nearby projections, as well as non shift-invariant covariance. These non-uniformities do not noticeably affect the resulting image quality if one reconstructs directly from the interpolated sinogram and thus periodic spline interpolation may be appropriate if the images are only to be inspected visually. If, however, one seeks to perform any sort of principled smoothing on the projections or quantitation on the reconstructed image that requires knowledge of the satistical properties of the projections, then the "stationarity"-preserving properties of CST/ZP interpolation may be the best choice.

Finally, the ability of CST and ZP interpolation to reduce the variance in the interpolated curve by the factor $(2K + 1)/N$ relative to the variance in the measured samples deserves comment. This is simply an implicit exploitation of the ability to achieve noise reduction through oversampling and filtering [12, Ch. 5], and the mechanism is easiest to appreciate in the case of zero-padding interpolation for stationary white noise. The DFT of Eq. 3 has N terms, corresponding to frequencies out to $k = \pm N/2$ for N even or $k = \pm(N - 1)/2$ for N odd. In Eq. 4, however, we see that all DFT components corresponding to frequencies $|k| > K$ are explicitly set equal to zero, because $g(x)$ is assumed to be bandlimited (at least approximately) to frequency K. However, if the measured samples are corrupted by white noise, then their DFT can be seen as the sum of the true DFT and the DFT of the noise process, which has contributions at all N frequency components. Zeroing out all but $2K + 1$ of these would thus be expected to reduce the noise magintude by the factor $(2K + 1)/N$.

V. References

[1] R. A. Brooks, G. H. Weiss, and A. J. Talbert, "A new approach to interpolation in computed tomography," *J. Comput. Assist. Tomogr.*, vol. 2, pp. 577–585, 1978.

[2] H. Stark, J. W. Woods, I. Paul, and R. Hingorani, "Direct Fourier reconstruction in computer tomography," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 29, pp. 237–245, 1981.

[3] H. Stark, "Sampling theorems in polar coordinates," *J. Opt. Soc. Am.*, vol. 69, pp. 1519–1525, 1979.

[4] H. Stark and M. Wengrovitz, "Comments and corrections on the use of polar sampling theorems in CT," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 31, pp. 1329–1331, 1983.

[5] R. W. Schafer and L. R. Rabiner, "A digital signal processing approach to interpolation," *Proc. IEEE*, vol. 61, pp. 692–702, 1973.

[6] K. P. Prasad and P. Satyanarayana, "Fast interpolation algorithm using FFT," *Electron. Lett.*, vol. 22, pp. 185–187, 1986.

[7] D. Fraser, "Interpolation by the FFT revisited—an experimental investigation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 665–675, 1989.

[8] T. J. Cavicchi, "DFT time-domain interpolation," *IEE Proceedings-F*, vol. 139, pp. 207–211, 1992.

[9] P. J. La Rivière and X. Pan, "Mathematical equivalence of zero-padding and circular sampling thorem interpolation with implications for direct Fourier image reconstruction," in *Proc. SPIE*, vol. 3338, pp. 1117–1126, 1998.

[10] H. Spath, *One-Dimensional Spline Interpolation Algorithms*. Wellesley, MA: A. K. Peters, 1993.

[11] C. H. Reinsch, "Smoothing by spline functions," *Numer. Math.*, vol. 10, pp. 177–183, 1967.

[12] R. J. Marks, *Introduction to Shannon Sampling and Interpolation Theory*. New York: Springer Verlag, 1991.

Comparison of angular interpolation approaches in few-view tomography using statistical hypothesis testing

P. J. La Rivière and X. Pan

Department of Radiology, The University of Chicago, Chicago, IL

ABSTRACT

In this work we examine the accuracy of four periodic interpolation methods—circular sampling theorem interpolatiom, zero-padding interpolation, periodic spline interpolation, and linear interpolation with periodic boundary conditions—for the task of interpolating additional projections in a few-view sinogram. We generated 100 different realizations each of two types of numerical phantom—Shepp-Logan and breast—by randomly choosing the parameters that specify their constituent ellipses. Corresponding sinograms of 128 bins x 1024 angles were computed analytically and subsampled to 16, 32, 64, 128, 256, and 512 views. Each subsampled sinogram was interpolated to 1024 views by each of the methods under consideration and the normalized root-mean-square-error (NRMSE) with respect to the true 1024-view sinogram computed. In addition, images were reconstructed from the interpolated sinograms by FBP and the NRMSE with respect to the true phantom computed. The non-parametric signed rank test was then used to assess the statistical significance of the pairwise differences in mean NRMSE among the interpolation methods for the various conditions: phantom family (Shepp-logan or breast), number of measured views (16, 32, 64, 128, 256, or 512), and endpoint (sinogram or image). Periodic spline interpolation was found to be superior to the others in a statistically significant way for virtually every condition.

Keywords: Few-view tomography, interpolation, image reconstruction, sampling, zero-padding, spline, circular sampling theorem

1. INTRODUCTION

In nuclear-medicine tomographic studies, there is generally a strong correlation between imaging time and image quality. Longer acquisitions lead to better measurement statistics and thus to less noisy, more quantitatively accurate reconstructed images. These gains are negated, however, if the long acquisition times lead to patient motion and thus to motion artifacts in the reconstructed images. Such artifacts are especially common, or expected to be, in studies such as cardiac single-photon emission computed tomography (SPECT) and the anticipated dependent-breast SPECT scintimammography[1,2] where older, relatively inflexible patients need to assume an uncomfortable position. In these studies, the ability actually to *reduce* current imaging times without severely compromising basic image quality would likely result in better overall imaging performance due to the reduction of motion artifacts. As an added benefit, the reduced imaging times would increase patient throughput in busy clinics.

For a constant patient dose, imaging time can be reduced by reducing the number of angular projections acquired, by reducing the amount of time spent acquiring each projection, or by some combination of the two. While the distinction between the two is admittedly blurred when using a continuous acquisition mode, we have found in preliminary studies that so long as certain minimal angular sampling requirements are met, having fewer angular projections with more counts leads to a higher ideal-observer signal-to-noise ratio in the reconstructed images than does having more angular projections with fewer counts, given the same total number of counts. In the case of a step-and-shoot acquisition protocol, which is always used for gated cardiac SPECT, reducing the number of angular views has the additional benefit of reducing the amount of deadtime spent moving the camera between views. In this paper, then, we focus on algorithms tailored to generate diagnostically useful images from a smaller number of angular views than is usually used while holding the number of counts per view constant; that is, we focus on algorithms for *few-view* tomography.

The minimum number of angular views required to produce an accurate, artifact-free tomographic reconstruction of a given object is dictated by two factors. First, the angular sampling of the object's sinogram must satisfy, at least approximately, the Nyquist sampling condition.[3] Absent this, any reconstruction is doomed to suffer from angular aliasing artifacts. Second, the number of angular samples must satisfy the reconstruction algorithm's implicit assumptions about the density of angular sampling. For instance, it is well known that images reconstructed from a small number of angular views by filtered backprojection (FBP) are degraded by prominent star-shaped artifacts.

These two conditions are not in general equivalent, as can be appreciated most keenly when considering the case of imaging a circularly symmetric object. In this instance, a single projection view is sufficient to satisfy the Nyquist sampling condition, while a FBP reconstruction from this single view would be an uninterpretable set of parallel streaks. Indeed, Brooks *et al.*[4] have shown that for a circularly symmetric object imaged with n bins per projection, a minimum of $\sim 1.1\pi n/4$ projections must be acquired over 180° to produce an essentially artifact-free reconstruction using FBP.

The Nyquist condition is the more fundamental of the two sampling requirements discussed above, because when it is satisfied by a number of samples less than the number required by the reconstruction algorithm, it is in principle possible to interpolate exactly the additional views needed. This is only strictly true in the absence of noise, of course, a condition that rarely obtains in emission tomography. We have discussed elsewhere the use of a principled smoothing technique based on non-parametric regression with an explicit Poisson statistical model to control noise in each projection prior to the interpolation of additional angular views.[5] We have also examined the noise properties of various interpolation methods when confronted with noise-corrupted samples.[6] In the present work, then, we wish simply to address the question of which interpolation approach is most accurate for the interpolation problems encountered in few-view tomography in the absence of noise, or when noise has been controlled in a previous step.

Because the sinogram is periodic in the angular dimension, the interpolation method chosen should rightly be periodic. In Section 2.1 we review four periodic interpolation methods being evaluated: circular sampling theorem interpolation, zero-padding interpolation, periodic spline interpolation, and linear interpolation with periodic boundary conditions. In Section 2.2 we discuss the statistical hypothesis testing approach we have taken to judging the relative accuracy of the approaches in the face of various few-view tomography interpolation tasks. In Section 3 we present the results of these evaluations, and finally, in Section 4 we present our conclusions about which interpolation method is best suited to the task encountered in few-view tomography.

Finally, it should be mentioned that in this paper we focus exclusively on few-view reconstruction involving FBP even though iterative reconstruction algorithms, such as algebraic reconstruction techniques and maximum likelihood expectation-maximization, should in principle fare better when reconstructing from a small number of views because they make no implicit assumptions about the completeness or continuity of the projection dataset. While an exploration of the performance of iterative algorithms in the face of few-view datasets will be the subject of later work, we felt it important to begin with an examination of FBP, which remains the most widely used algorithm in clinical settings.

2. METHODS

2.1. Periodic Interpolation Methods

2.1.1. Circular sampling theorem and zero-padding interpolation

Circular sampling theorem (CST) interpolation is a special case of Whittaker-Shannon (W-S) sinc interpolation that applies to periodic functions.[7,8] Consider a periodic function $g(x)$ that has period X and which is bandlimited to frequency K (i.e., the coefficients of expansion a_k of the function's Fourier series satisfy $a_k = 0$ for $|k| > K$). Given $N \geq 2K + 1$ samples of $g(x)$ taken at points $x_n = nX/N$ ($n = 0, \ldots, N - 1$) evenly spaced over one period, the CST states[8] that $g(x)$ can be interpolated exactly by use of

$$\widehat{g}(x) = \sum_{n=0}^{N-1} g(x_n)\sigma_N(x - x_n), \tag{1}$$

where

$$\sigma_N(x) = \sin[(2K + 1)\pi x/X]/N\sin(\pi x/X). \tag{2}$$

If the Nyquist condition is not satisfied, that is, if $g(x)$ is not truly bandlimited to frequency K or if $N < 2K + 1$, Eqs. (1) and (2) no longer represent exact interpolation, but they remain mathematically meaningful and, often, practically useful. For instance, if the spectral components beyond frequency K are negligibly small but not exactly zero, interpolating with Eqs. (1) and (2) remains very accurate.

A second periodic interpolation approach, zero-padding (ZP) interpolation, involves extending the discrete Fourier transform (DFT) of a finite sequence with zeroes and taking an inverse DFT to generate a more densely sampled

version of the original sequence with values interpolated at intermediate positions between the original measured samples.[9-12] Specifically, one begins by taking the DFT of the sequence $g(x_n)$, which is given by

$$c_k = \frac{1}{N} \sum_{n=0}^{N-1} g(x_n) \exp(-j2\pi nk/N), \tag{3}$$

for $k = 0, \ldots, N-1$, where $j = \sqrt{-1}$. Zero-padding involves the creation of a new sequence $d_{k'}$, having $L = PN$ elements (where P is an integer). If $g(x)$ is assumed to be bandlimited to frequency K and if $N \geq 2K + 1$, the sequence $d_{k'}$ is defined as follows:

$$d_{k'} = \begin{cases} c_{k'} & k' = 0, \ldots, K \\ 0 & k' = K+1, \ldots, L-K-1 \\ c_{k'-L+N} & k' = L-K, \ldots, L-1. \end{cases} \tag{4}$$

A more densely sampled sequence $\widehat{g}(x_l)$, where $x_l = lX/L$ ($l = 0, \ldots, L-1$), is now obtained by taking the inverse DFT of the sequence $d_{k'}$,

$$\widehat{g}(x_l) = \sum_{k'=0}^{L-1} d_{k'} \exp(j2\pi lk'/L), \tag{5}$$

for $l = 0, \ldots, L-1$. ZP interpolation is generally viewed as a somewhat crude interpolation approach, but this reputation is undeserved. It can be shown that ZP interpolation is equivalent to CST interpolation, in that the spatial-domain interpolation function corresponding to the ZP operation in frequency space is just $\sigma_N(x)$ given in Eq. (2).[13] That is, it can be shown that

$$\widehat{g}(x_l) = \sum_{n=0}^{N-1} g(x_n)\sigma_N(x_l - x_n), \tag{6}$$

where $\sigma_N(x)$ is given by Eq. (2). This equivalence holds regardless of whether the Nyquist condition is satisfied. If it is satisfied, then ZP interpolation, like CST interpolation, is exact. Obviously, the CST interpolation formula can be used to estimate $g(x)$ at any arbitrary point x whereas zero-padding interpolation is constrained to interpolate onto a fixed, equispaced grid P times denser than the original samples. However, since ZP yields a continuous interpolated curve in the limit as $P \to \infty$, we shall treat these two approaches as one in the subsequent analysis, using the continuous form of Eq. (1) to represent them both. Given this equivalence, ZP is in general to be preferred in practice because it is considerably more computationally efficient than the CST. Whenever using ZP/CST below, we take K to be $N/2 - 1$; that is, we assume that the bandlimit of the function is equal to (or higher than) than the Nyquist frequency of the samples.

2.1.2. Periodic spline interpolation

Splines are piecewise cubic polynomials that are continuous up to and including the second derivative at the joints between pieces.[14,15] Periodic splines are further constrained to satisfy periodic boundary conditions. A spline $\widehat{g}(x)$ can be represented by

$$\widehat{g}(x) = a_n + b_n(x - x_n) + c_n(x - x_n)^2 + d_n(x - x_n)^3, \tag{7}$$

for $x \in [x_n, x_{n+1}]$, where the x_n are the abscissas at which the data is measured and $n = 0, \ldots, N - 1$. Fitting a spline is thus tantamount to finding the coefficients a_n, b_n, c_n, and d_n subject to interpolation, continuity, and boundary conditions. A number of efficient algorithms exist for doing so.[14]

2.1.3. Linear interpolation with periodic boundary conditions

The last approach to be considered is also the simplest: linear interpolation with periodic boundary conditions. In this case the interpolation takes the form

$$\widehat{g}(x) = (1 - w_n(x))g(x_n) + w_n(x)g(x_{n+1}), \tag{8}$$

for $x \in [x_n, x_{n+1}]$, where the x_n are the abscissas at which the data is measured, $n = 0, \ldots, N - 1$, and $w_n(x) = (x - x_n)/(x_{n+1} - x_n)$. The peridocity condition enters when $n = N - 1$, in which case $g(x_N)$ is taken to be equal to $g(x_o)$.
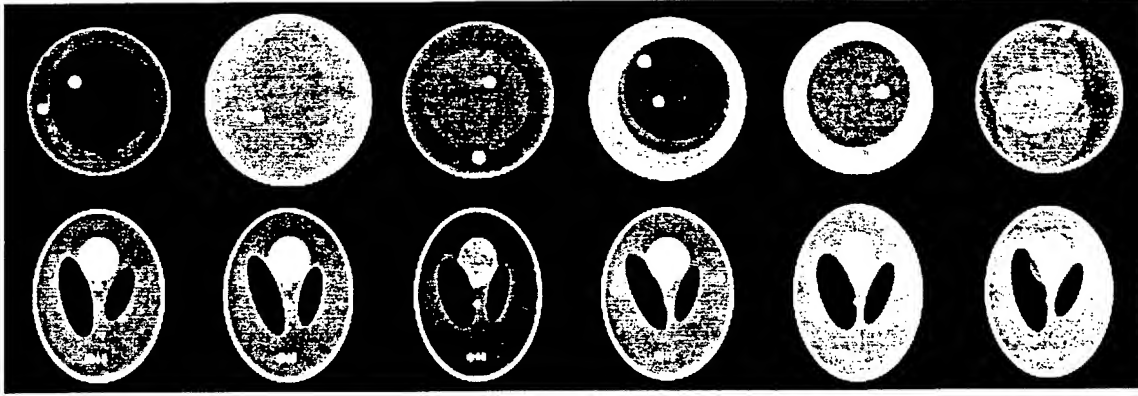
Figure 1. Six typical realizations of the breast (upper) and Shepp-Logan (lower) types of phantoms whose sinograms were used in the interpolation accuracy studies. The realizations were generated by choosing the parameters governing the phantoms' constituent ellipses at random according to predetermined probability laws.

2.2. Evaluation of Accuracy Using Statistical Hypothesis Testing

Various theoretical claims can be made about the accuracy of the interpolation methods being considered, each depending on the degree to which the data satisfy the assumptions underlying the method. For instance, linear interpolation is exact if the true function is piecewise linear between the measured samples and quite accurate if the function is sampled densely enough that it is nearly linear between measured samples. The same holds for spline interpolation and piecewise cubic, or nearly so, functions. CST and ZP interpolations, as noted above, are exact for periodic functions that are bandlimited to the Nyquist frequency of the samples. Another approach to the evaluation of interpolation methods involves examining the Fourier transforms of the methods' interpolation kernels to see how well they approximate the rectangular transform of the theoretically exact sinc kernel. This again implicitly assumes that the function is bandlimited to the Nyquist frequency of the samples, or, at least, that violations of the bandlimited assumption only negligibly compromise the accuracy of the approach.

Real data rarely satisfy any of these assumptions exactly, and in practice we have found that mild violations, particularly of the bandlimited assumption, can lead to undesirable errors. We wished to judge the methods being considered, then, not on the basis of theoretical claims but empirically in the context of the task of interest: the interpolation of additional angular views in a sinogram. To do so, we required numerical phantoms whose projections could be calculated analytically. In this way, we could generate a sinogram with a reduced number of views, interpolate to a larger number of views, and then compare quantitatively the interpolated sinogram to the exact sinogram for the corresponding number of views. Fortunately, it is possible to compute the projections an ellipse of arbitrary size, position, and orientation exactly,[16] allowing much freedom in the design of objects whose sinograms were to be interpolated. We focused on two types of phantoms: the familiar Shepp-Logan brain phantom[17] and a simulated breast phantom consisting of a large outer circle, some slightly smaller background ellipses, and small circular lesions.

Simply comparing the success of the approaches in interpolating a single sinogram each for one or two canonical phantoms would provide more anecdotal than genuinely rigorous evidence on which to base the choice of interpolation method for few-view tomography. The outcome could depend as much on numerical happenstance as on the genuine strengths of the approaches. Instead, we generated 100 different "realizations" of each of the two types of numerical phantom by choosing the parameters specifying the constituent ellipses of each type to vary according to predetermined probability laws. Some typical phantom realizations are illustrated in Figure 1.

For each realization of the phantom parameters we generated a parallel-beam sinogram of 128 bins and 1024 angular views. We then subsampled the sinogram to generate sinograms of 16, 32, 64, 128, 256, and 512 views. Each of these subsampled sinograms was interpolated to 1024 views by applying each of the three periodic interpolation approaches discussed in turn. The resulting interpolated sinograms were then compared to the analytically computed 1024-view sinogram and the normalized root-mean-square error (NRMSE) between the two computed. This error measure is defined as

$$d = \left(\sum_{i=1}^{R} \sum_{i=1}^{R} (t_{ij} - r_{ij})^2 / \sum_{i=1}^{R} \sum_{i=1}^{R} (t_{ij} - \bar{t})^2 \right)^{1/2} , \qquad (9)$$

	ZP vs. Spline			Spline vs. Linear			Linear vs. ZP		
# views	Mean Diff.	W_+	p-value	Mean Diff.	W_+	p-value	Mean Diff.	W_+	p-value
16	1.09×10^{-3}	29.0	0.0000	4.96×10^{-4}	1234.0	0.0000	-1.58×10^{-3}	4572.0	0.0000
32	6.98×10^{-4}	12.0	0.0000	-4.40×10^{-4}	4286.0	0.0000	-2.58×10^{-4}	2946.0	0.0739
64	3.10×10^{-4}	0.0	0.0000	-3.91×10^{-4}	5024.0	0.0000	8.19×10^{-5}	1882.0	0.0135
128	1.72×10^{-4}	4.0	0.0000	-1.51×10^{-4}	5003.0	0.0000	-2.0×10^{-5}	3040.0	0.0383
256	9.06×10^{-5}	0.0	0.0000	-6.86×10^{-5}	4890.0	0.0000	-2.2×10^{-5}	3664.0	0.0000
512	4.83×10^{-5}	1.0	0.0000	-2.95×10^{-5}	4835.0	0.0000	-1.88×10^{-5}	4239.0	0.0000

Table 1. Pairwise mean breast phantom sinogram NRMSE differences, test statistics, and p-values for the three interpolation methods.

where t_{ij} and r_{ij} represent the pixel values of the pixel in the ith row and jth column of the true and reconstructed RxR images, respectively, and $\bar{t}$ denotes the average pixel value in the true image. In order to see how the sinogram interpolation error affected the accuracy of reconstructed images, we also used FBP to reconstruct an image from each sinogram interpolated by each of the three approaches, and then calculated the NRMSE with respect to the true phantom.

The result of this simulation was a set of 100 NRMSE values for each combination of phantom family (Shepp-logan or breast), interpolation method (CST/ZP, spline, or linear), number of measured views (16, 32, 64, 128, 256, or 512), and endpoint (sinogram or image). This data lent itself naturally to a statistical hypothesis testing analysis seeking to evaluate for each task (phantom family, number of measured angles, and endpoint) the null hypothesis that, considered pairwise, the three interpolation methods yielded the same NRMSE. Refuting the null hypothesis would allow the methods to be ranked for each task.

Because the interpolation methods were applied in parallel to the same realizations of the phantom, a paired test must be used to evaluate the null hypotheses. The familiar paired t-test, in which the differences between paired samples are compared to cutoffs of the t-distribution, implicitly assumes that the differences are a sample from a normal distribution. As exploratory data analysis indicated that the distribution of differences was not evidently normal in the present case, we turned instead to the nonparametric signed rank (SR) test for paired samples.[18] In the SR test, the absolute values of the differences are ranked (with appropriate handling of ties), the signs of the differences applied to the corresponding ranks, and the sum W_+ of those ranks having positive signs computed. The distribution of W_+ under the null hypothesis that the two conditions produce the same values can be easily computed, and extreme values of W_+ signal that one condition tends to produce larger value than the other. Finally, the p-value (the probability that a W_+ as extreme or more extreme than the one found could arise under the null hypothesis) of each W_+ can be computed from this distribution as well.

3. RESULTS

The results for the interpolation of the family of breast phantoms are depicted in Figure 2. For each interpolation method, Figure 2(a) plots the mean sinogram NRMSE for the 100 realizations versus the number of measured projections. A higher NRMSE corresponds to worse accuracy. Because the curves in Figure 2(a) are difficult to distinguish, particularly for large numbers of measured projections, the NRMSE relative to the periodic spline NRMSE at each number of measured projections is plotted in Figure 2(b). The corresponding curves for reconstructed image NRMSE are depicted in Figures 2(c) and 2(d). Tables 1 and 2 list the mean difference between each pair of methods for each number of measured projections as well as the corresponding value of the SR statistic W_+ and its two-sided p-value. A low p-value indicates that the mean difference between the two methods is statistically significant, and we adopt throughout a stringent 0.01 cutoff for statistical significance.

A few trends common to all three interpolation methods can be gleaned from Figure 2. For one, the accuracy of all three approaches, as measured by the sinogram NRMSE, improves rapidly as the number of measured projections increases. This is not surprising—the assumptions underlying all three approaches are better satisfied as the number of samples increases—but it does offer an interesting contrast to the trend seen in reconstructed image NRMSE. These NRMSE values level off beyond 64 samples, with the change in overall NRMSE from one number of measured
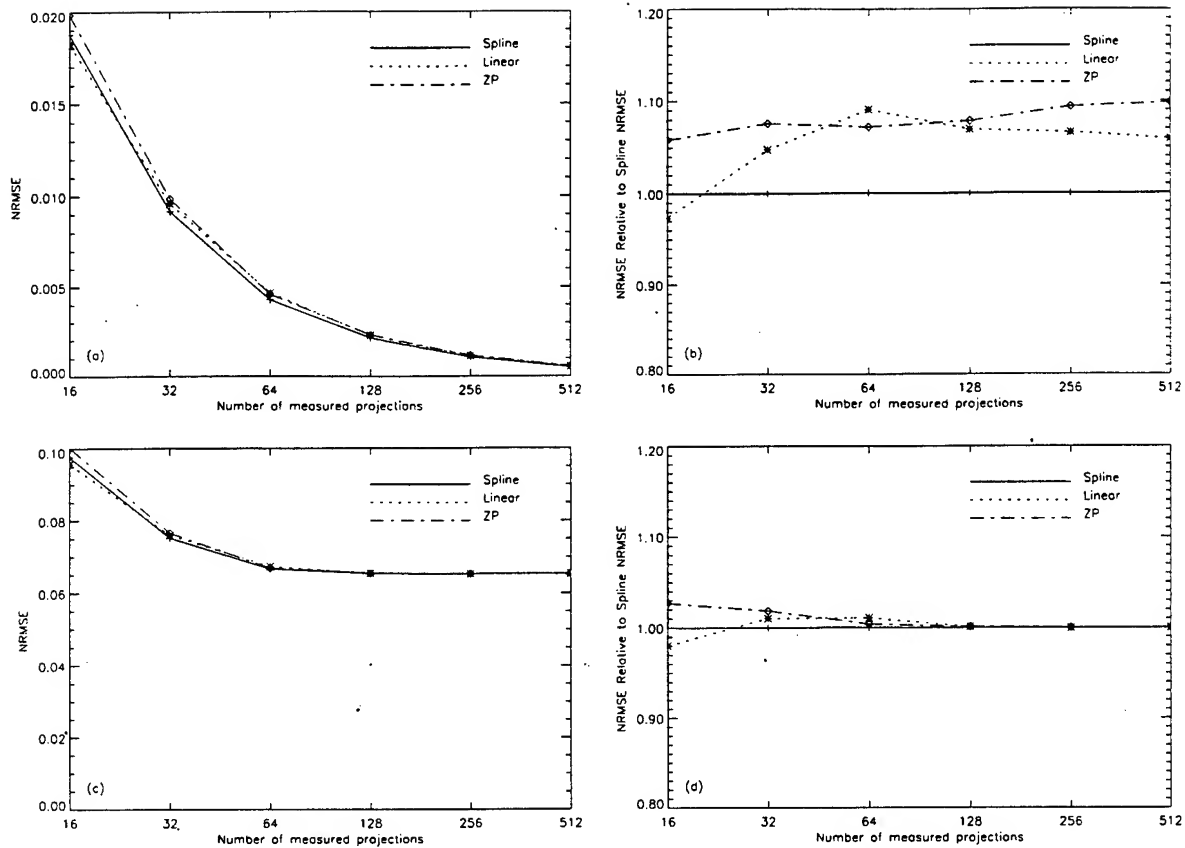
Figure 2. NRMSE and relative NRMSE plots for the family of breast phantoms. (a) Plot of sinogram NRMSE vs. number of measured projections for each interpolation method. (b) Plot of sinogram NRMSE relative to spline sinogram NRMSE vs. number of measured projections for each interpolation method. (c) Plot of reconstructed image NRMSE vs. number of measured projections for each interpolation method. (d) Plot of reconstructed image NRMSE relative to spline image NRMSE vs. number of measured projections for each interpolation method.

	ZP vs. Spline			Spline vs. Linear			Linear vs. ZP		
# views	Mean Diff.	W_+	p-value	Mean Diff.	W_+	p-value	Mean Diff.	W_+	p-value
16	2.65×10^{-3}	88.0	0.0000	1.91×10^{-3}	966.0	0.0000	-4.56×10^{-3}	4520.0	0.0000
32	1.40×10^{-3}	147.0	0.0000	-7.76×10^{-4}	3771.0	0.0000	-6.26×10^{-4}	2659.0	0.3225
64	2.54×10^{-4}	138.0	0.0000	-7.03×10^{-4}	4237.0	0.0000	4.49×10^{-4}	1444.0	0.0001
128	6.74×10^{-5}	3.0	0.0000	-5.33×10^{-5}	2281.0	0.2008	-1.41×10^{-5}	3227.0	0.0079
256	3.30×10^{-6}	75.0	0.0000	2.69×10^{-5}	1005.0	0.0000	-3.02×10^{-5}	4138.0	0.0000
512	2.00×10^{-7}	1289.5	0.0000	7.90×10^{-6}	509.0	0.0000	-8.00×10^{-6}	4466.5	0.0000

Table 2. Pairwise mean breast phantom image NRMSE differences, test statistics, and p-values for the three interpolation methods.

projections to the next being comparable to the differences among the interpolation method for a given number of measured projections. This leveling off is of course due to the fact that errors associated with other inaccuracies in the image reconstruction process, such as finite radial sampling, come to overwhelm the error due to finite angular sampling as the number of angular samples grows large.

As for the relative performance of the three interpolation approaches, when comparing sinogram NRMSEs, periodic spline interpolation is seen to be superior to the other two in a statistically significant way for all but the
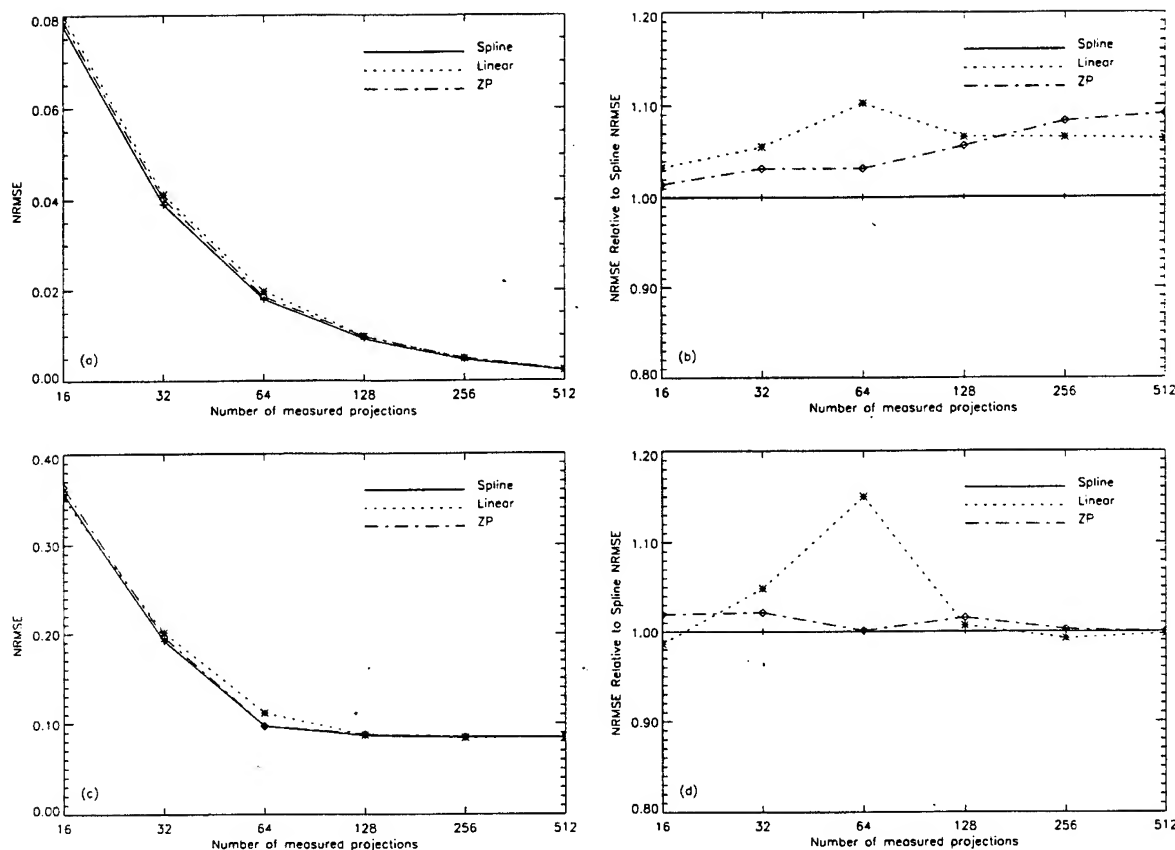
Figure 3. NRMSE and relative NRMSE plots for the family of Shepp-Logan phantoms. (a) Plot of sinogram NRMSE vs. number of measured projections for each interpolation method. (b) Plot of sinogram NRMSE relative to spline sinogram NRMSE vs. number of measured projections for each interpolation method. (c) Plot of reconstructed image NRMSE vs. number of measured projections for each interpolation method. (d) Plot of reconstructed image NRMSE relative to spline image NRMSE vs. number of measured projections for each interpolation method.

smallest number of samples—16—where linear interpolation performs best. Linear interpolation is seen to be better than or statistically indistinguishable from ZP interpolation for all numbers of measured projections. Similar trends are observed in comparing image NRMSEs. Spline outperforms ZP interpolation in every case, and is better than linear interpolation for moderate numbers of measured projections, but slightly worse or indistiguishable for very large numbers of samples. Linear is again seen to be better than or indistinguishable from ZP for all numbers of measured projections except 64. It may at first seem paradoxical that comparing image NRMSEs does not always yield the same relative performance as sinogram NRMSEs. Shouldn't more accurate sinograms necessarily produce more accurate images? The resolution lies in the fact that NRMSEs are global error measures obtained by summing over local errors. It is the local errors that propagate directly into the reconstructed images and these may combine or cancel differently depending on their relative positions in the sinogram. Thus the geography of the local sinogram errors, as well as their sum, influences the global error in the reconstruction, and so a sinogram with a higher NRMSE than another can in principle lead to a reconstructed image with a lower NRMSE than the other.

The corresponding results for the family of Shepp-Logan phantoms are illustrated in Figure 3 and Tables 3 and 4. The same trends are observed with regard to the overall performance of the three approaches. The sinogram NRMSEs fall rapidly with increasing number of measured projections, while the reconstructed image NRMSEs level off rapidly beyond 64 projections. While these trends are as before, it should be pointed out that the magnitude of the NRMSEs for both sinograms and reconstructed images is significantly higher for the Shepp-Logan type phantoms than it was for the breast phantoms, owing to the fact that the Shepp-Logan phantoms are more complex and their sinograms more variable in the angular direction than the breast phantoms.

# views	ZP vs. Spline			Spline vs. Linear			Linear vs. ZP		
	Mean Diff.	W_+	p-value	Mean Diff.	W_+	p-value	Mean Diff.	W_+	p-value
16	1.12×10^{-3}	0.0	0.0000	-2.58×10^{-3}	5045.0	0.0000	1.45×10^{-3}	575.0	0.0000
32	1.24×10^{-3}	53.0	0.0000	-2.17×10^{-3}	5032.0	0.0000	9.18×10^{-4}	1530.0	0.0003
64	5.72×10^{-4}	747.0	0.0000	-1.83×10^{-3}	4503.0	0.0000	1.26×10^{-3}	1252.0	0.0000
128	5.27×10^{-4}	9.0	0.0000	-6.19×10^{-4}	4945.0	0.0000	9.20×10^{-5}	2136.0	0.0905
256	3.82×10^{-4}	1.0	0.0000	-3.06×10^{-4}	4845.0	0.0000	-7.58×10^{-5}	3102.0	0.0236
512	1.94×10^{-4}	0.0	0.0000	-1.39×10^{-4}	4722.0	0.0000	-5.52×10^{-5}	3203.0	0.0099

Table 3. Pairwise mean Shepp-Logan phantom sinogram NRMSE differences, test statistics, and p-values for the three interpolation methods.

# views	ZP vs. Spline			Spline vs. Linear			Linear vs. ZP		
	Mean Diff.	W_+	p-value	Mean Diff.	W_+	p-value	Mean Diff.	W_+	p-value
16	6.91×10^{-3}	0.0	0.0000	4.88×10^{-3}	244.0	0.0000	-1.18×10^{-2}	5050.0	0.0000
32	4.10×10^{-3}	177.0	0.0000	-9.31×10^{-3}	5050.0	0.0000	5.21×10^{-3}	968.0	0.0003
64	1.07×10^{-4}	2454.0	0.4036	-1.46×10^{-2}	5050.0	0.0000	1.44×10^{-2}	0.0	0.0000
128	1.36×10^{-3}	0.0	0.0000	-5.95×10^{-4}	3909.0	0.0000	-7.68×10^{-4}	4153.0	0.0905
256	2.19×10^{-4}	0 .0	0.0000	6.57×10^{-4}	0.0	0.0000	-8.77×10^{-4}	5050.0	0.0236
512	9.00×10^{-6}	1.0	0.0000	2.12×10^{-4}	0.0	0.0000	-2.21×10^{-4}	5050.0	0.0099

Table 4. Pairwise mean Shepp-Logan phantom image NRMSE differences, test statistics, and p-values for the three interpolation methods.

As for the relative performance of the interpolation methods for the Shepp-Logan phantoms, the results are similar but not identical to those of the breast phantom. When comparing sinogram NRMSEs, periodic spline interpolation is seen to be better than ZP or linear interpolation in a statistically significant way for all numbers of measured projections. Linear interpolation is better than ZP for larger numbers of projections (256 or 512), worse for small numbers (16, 32, and 64), and indistingushable for 128 projections. When comparing reconstructed image NRMSEs, spline is still better than ZP for every number of measured projections, and, as before, better than linear interpolation for moderate numbers of projections (32, 64, and 128) while worse for extreme numbers (16, 256, and 512). Linear is better than zero-padding for extreme numbers of measured projections (16, 128, 256, and 512) and worse for moderate numbers (32 and 64).

4. DISCUSSION AND CONCLUSIONS

We have presented a study of the empirical accuracy of three different interpolation approaches for the task of interpolating additional angular views in a sinogram. For two types of numerical phantom—Shepp-Logan and breast—we generated 100 different realizations by randomly choosing the parameters that specify their constituent ellipses. Sinograms of 128 bins x 1024 angles were computed analytically and subsampled to 16, 32, 64, 128, 256, and 512 views. Each subsampled sinogram was interpolated to 1024 views by each of the three methods under consideration—CST/ZP, periodic spline, and linear interpolation—and the NRMSE compared to the true 1024-view sinogram computed. The non-parametric signed rank test was then used to assess the statistical significance of the pairwise differences in mean NRMSE among the interpolation methods for the various conditions: phantom family (Shepp-logan or breast), number of measured views (16, 32, 64, 128, 256, or 512), and endpoint (sinogram or image).

Periodic spline interpolation was found to be superior in a statistically significant way to CST/ZP and linear interpolation for virtually every condition. It was superior to CST/ZP in every instance, and fell short of linear interpolation only in the case of a very small number of measured angles (16) for breast sinogram NRMSE and for extremely large or small numbers of angles for reconstructed image NRMSE for both types of phantoms. It was certainly always superior for interpolation from the moderate number of angles—32 and 64—likely to be relevant to few-view tomography. The strong performance of spline interpolation can be attributed to the combination of its

great flexibility and relatively local response. That is, while splines can capture subtle features that frequently elude piecewise linear curves, they do not allow sharp edges that they are unable to capture faithfully to influence unduly the accuracy of the curve at distant points. CST/ZP interpolation, on the other hand, while blessed with considerably more flexibility than linear interpolation, tends to suffer from widespread ringing artifacts (Gibbs phenomena)[19] when sharp edges cause a violation, however mild, of the bandlimited assumption on which it is built.

Indeed, the relatively poor performance of CST/ZP interpolation was one of the surprises in this work, and careful examination of individual curves interpolated by this method indicated that Gibbs artifacts were indeed the principal cause of the poor performance. Ironically, while the severity of the ringing increases as the number of measured samples decreases, zero-padding performed relatively well in this range compared to linear interpolation (though not to spline), and generally fell behind linear only for denser sampling. This is likely because at this low sampling density the piecewise linear approximation is so poor over most of the function being interpolated that CST/ZP, with its more flexible curves, performs better despite the ringing. The ringing only becomes a factor as both methods perform similarly over the slowly varying parts of the curve, with linear interpolation's local response around edges then being preferable to the ringing introduced by CST/ZP. Attempts to mitigate the ringing using anti-oscillation filters such as Lanczos's sigma filters[20] met with some success, but the smoothing caused by these filters degraded the NRMSE performance at least as much as the ringing had. In short, unless the few-view data is known to be precisely bandlimited, CST/ZP interpolation is not recommended.

This phenomenon of non-local response may also explain linear interpolation's suprising success relative to spline interpolation for large numbers (256 or 512) of measured projections. With sampling of such density, the piecewise linear approximation is of course very good over most of the curve, and those few points where the approximation is not very good contribute little to the total NRMSE. The accuracy of spline interpolation, on the other hand, is likely to be comparable to linear interpolation over most of the interpolated points, but may suffer more widely the effects of an edge because the spline interpolation kernel is considerably wider than that of linear interpolation. Nonetheless, spline interpolation was seen to be clearly superior for both simple and complex objects in precisely the angular sampling range (32-64) of interest to few-view tomography.

ACKNOWLEDGMENTS

REFERENCES

1. P. J. La Rivière, X. Pan, and B. C. Penney, "Ideal-observer analysis of lesion detectability in planar, conventional SPECT, and dedicated SPECT scintimammography using effective multi-dimensional smoothing," *IEEE Trans. Nucl. Sci.* **45**, pp. 1273–1279, 1998.

2. H. Wang, C. Scarfone, K. L. Greer, R. E. Coleman, and R. J. Jaszczak, "Prone breast tumor imaging using vertical axis-of-rotation SPECT systems: An initial study," *IEEE Trans. Nucl. Sci.* **44**, pp. 1271–1276, 1997.

3. R. J. Marks, *Introduction to Shannon Sampling and Interpolation Theory*, Springer Verlag, New York, 1991.

4. R. A. Brooks, G. H. Weiss, and A. J. Talbert, "A new approach to interpolation in computed tomography," *J. Comput. Assist. Tomogr.* **2**, pp. 577–585, 1978.

5. P. J. La Rivière and X. Pan, "Few-view tomography using interpolating and smoothing splines with implications for cardiac spect," in *Proc. IEEE Nucl. Sci. Symp. and Med. Imag. Conf.*, 1999. (To appear).

6. P. J. La Rivière and X. Pan, "Noise properties of periodic interpolation methods with implications for few-view tomography," in *Proc. IEEE Nucl. Sci. Symp. and Med. Imag. Conf.*, 1999. (To appear).

7. H. Stark, "Sampling theorems in polar coordinates," *J. Opt. Soc. Am.* **69**, pp. 1519–1525, 1979.

8. H. Stark and M. Wengrovitz, "Comments and corrections on the use of polar sampling theorems in CT," *IEEE Trans. Acoust., Speech, Signal Processing* **31**, pp. 1329–1331, 1983.

9. R. W. Schafer and L. R. Rabiner, "A digital signal processing approach to interpolation," *Proc. IEEE* **61**, pp. 692–702, 1973.

10. K. P. Prasad and P. Satyanarayana, "Fast interpolation algorithm using FFT," *Electron. Lett.* **22**, pp. 185–187, 1986.

11. D. Fraser, "Interpolation by the FFT revisited—an experimental investigation," *IEEE Trans. Acoust., Speech, Signal Processing* **37**, pp. 665–675, 1989.

12. T. J. Cavicchi, "DFT time-domain interpolation," *IEE Proceedings-F* **139**, pp. 207–211, 1992.

13. P. J. La Rivière and X. Pan, "Mathematical equivalence of zero-padding and circular sampling thorem interpolation with implications for direct Fourier image reconstruction," in *Proc. SPIE*, vol. 3338, pp. 1117–1126, 1998.

14. H. Spath, *One-Dimensional Spline Interpolation Algorithms*, A. K. Peters, Wellesley, MA, 1993.

15. C. H. Reinsch, "Smoothing by spline functions," *Numer. Math.* **10**, pp. 177–183, 1967.

16. A. Rosenfeld and A. C. Kak, *Digital Picture Processing*, Academic Press, New York, 1982.

17. L. A. Shepp and B. F. Logan, "The Fourier reconstruction of a head section," *IEEE Trans. Nucl. Sci.* **21**, pp. 21–43, 1974.

18. J. A. Rice, *Mathematical Statistics and Data Analysis*, Wadsworth, Belmont, CA, 1994.

19. E. O. Brigham, *The Fast Fourier Transform and Its Applications*, Prentice Hall, Englewood Cliffs, 1988.

20. C. Lanczos, *Applied Analysis*, Dover, Mineola, NY, 1988.

Improved Detectability of Malignant Lesions in SPECT Scintimammography Using Effective Multi-Dimensional Smoothing

P.J. La Rivière, X. Pan, B.C. Penney, and C.-T. Chen

Department of Radiology, The University of Chicago, Chicago, IL, 60637

Abstract

Scintimammography, a nuclear-medicine imaging technique that relies on the preferential uptake of Tc-99m-sestamibi and other radionuclides in breast malignancies, has the potential to provide differentiation of mammographically suspicious lesions as well as outright detection of malignancies in women with radiographically dense breasts. In this work we use the ideal-observer framework to quantify the detectability of a 1-cm lesion using three different imaging geometries: the planar technique that is the current clinical standard, conventional single-photon emission computed tomography (SPECT), in which the scintillation cameras rotate around the entire torso, and dedicated breast SPECT, in which the cameras rotate around the breast alone. We also introduce an adaptive smoothing technique for the processing of planar images and of sinograms prior to reconstruction that exploits Fourier transforms to achieve effective multidimensional smoothing at a reasonable computational cost.

For the detection of a 1-cm lesion with a clinically typical 6:1 tumor-background ratio, we find ideal-observer signal-to-noise ratios (SNR) of 6.5 for planar imaging without processing and 6.7 after applying this adaptive smoothing, 3.5 for conventional SPECT rising to 6.1 with adaptive smoothing, and finally 9.6 for the dedicated geometry rising to 11.4 with adaptive smoothing. The results suggest that the dedicated breast SPECT geometry is the most effective of the three, and that the adaptive, two-dimensional (2D) smoothing enhances idealized lesion detectability.

I. INTRODUCTION

In recent years, researchers have developed and studied a nuclear-medicine test with the potential to provide relatively low-cost, minimally invasive differentiation of breast abnormalities identified by physical examination or mammography [1,2]. Known as scintimammography, the test relies on the preferential uptake of Tc-99m-sestamibi or other radionuclides such as Tl-201, Tc-99m-tetrofosmin, or Tc-99m-MDP in breast malignancies as compared to normal breast tissue or benign abnormalities. Indeed, one study has shown that typical *in vivo* tumor-background concentration ratios of Tc-99m-Sestamibi are 5.64±3.06 [3]. This focal uptake can be imaged in a number of ways, though the most widely used clinical protocol involves acquiring one or two planar views while the patient lies prone [4]. The imaging time is typically 10-15 minutes per view. Numerous clinical studies with histological follow-up have been performed using this or a similar protocol, reporting sensitivities of 83-96%, and specificities of 66-100% when using Tc-99m-Sestamibi to image mammographically suspicious lesions [1,2, and references therein]. In addition to differentiating breast abnormalities detected by other means, scintimammography may also have a role in the detection of breast malignancies in patients with radiographically dense breasts for whom screen-film mammograms are often difficult to interpret [2].

A few of these studies of scintimammography have also examined the role of conventional SPECT (where the patient lies supine and the camera circles the torso) in detecting focal uptake of Tc-99m-sestamibi and have found comparable but not generally improved sensitivities as compared to planar techniques, along with substantially lower specificities [5-8]. Wang *et al.* [9] speculated that this surprisingly poor performance was due to substantial attenuation of photons emitted in the breast by the torso as well as to the presence of scatter from organs such as the heart and liver. The poor performance of conventional SPECT may also be related to the inferior resolution of conventional SPECT as compared to planar techniques in this situation, due to the fact that the scintillation cameras are on average further away from the breast in the conventional SPECT geometry. Wang *et al.* also investigated a geometry that they called vertical axis-of-rotation SPECT and that we call dedicated breast SPECT, in which the scintillation cameras are assumed to rotate around one breast alone. This geometry eliminates the effect of attenuation by the thorax and, with proper shielding, the effect of scatter from the thorax. Moreover, the small radius of rotation offers improved resolution and sensitivity. In phantom studies, Wang *et al.* found that with this dedicated geometry they were able to detect a breast lesion with an outer diameter of 1 cm and a 6:1 lesion-to-background concentration ratio that was not detectable in either conventional SPECT or planar studies with the same total imaging time.

The present work examines quantitatively the question of lesion detectability in these three different geometries using the so-called ideal-observer framework to calculate signal-to-noise ratios as a function of lesion concentration for the different geometries, and, in the case of the tomographic geometries, for different reconstruction filters. It also introduces and applies an adaptive smoothing technique for the processing of planar images or sinograms that exploits Fourier transforms to achieve effective multidimensional smoothing at a reasonable computational cost.

II. METHODS

A. Effective multi-dimensional smoothing

The projection data acquired by tomographic nuclear-medicine imaging devices is invariably contaminated by noise, which is propagated and often amplified by the

reconstruction algorithm. In filtered backprojection (FBP), the most computationally efficient and most commonly used of these algorithms, noise control is usually achieved through linear filtering, either of the projections prior to reconstruction or of the reconstructed image itself. In linear filtering, the degree of smoothing applied to the data does not depend on the data, but is rather fixed *a priori*, and this same degree of smoothing is applied to the entire dataset. In contrast, non-linear, adaptive smoothing methods, such as the generalized cross-validation method to be discussed below, determine the degree of smoothing to be applied to various specified subsets of the data from the statistics of each such subset and the degree of smoothing can vary from subset to subset. For instance, in 2D image reconstruction, the data at each projection angle could be smoothed differently, with the degree of smoothing determined from the statistics of the data at each angle.

When smoothing for 2D image reconstruction, one would in principle like to exploit the statistical correlations between different projections as well as those within a given projection. A truly 2D adaptive smoothing operation of this type can be computationally expensive and difficult to implement [10]. However, by exploiting the properties of the Fourier transform, one can achieve *effective* 2D smoothing at the cost of a series of one-dimensional (1D) smoothing operations. Consider a 2D discrete sinogram $p(\xi_i, \theta_j)$, where ξ_i refers to the ith projection bin *(i=1,...,N)* and θ_j the jth projection angle *(j=1,...M)*. It can be shown [11] that the following sequence of operations is equivalent to an adaptive, 2D smoothing of the sinogram:

1. Take a 1D discrete Fourier transform of the sinogram with respect to the projection angle θ_j; the result can be viewed as a set of complex 1D functions of the untransformed variable ξ_i, each labeled by an angular frequency index k, i.e. $P_k(\xi_i)$.

2. Perform an adaptive 1D smoothing of the real and imaginary parts of each of these M functions of ξ_i, yielding M discrete smoothed functions $P_k^s(\xi_i)$, where the superscript s stands for smoothed.

3. Perform an inverse 1D Fourier transform of $P_k^s(\xi_i)$ with respect to the angular frequency k to recover $p^s(\xi_i, \theta_j)$.

The adaptive 1D smoothing we perform on each of the M functions $P_k(\xi_i)$ is known as penalized least-squares smoothing [12,13], and involves fitting the discrete data with a continuous smoothing curve $P_k^s(\xi)$ that minimizes the functional

$$S\{P_k^s(\xi)\} = \sum_{i=1}^{N} \left[P_k(\xi_i) - P_k^s(\xi_i) \right]^2 + \alpha \int_0^{\xi_{max}} [P_k^{s''}(\xi)]^2 d\xi, \quad (1)$$

where ξ is a continuous variable representing the position along a given projection, ξ_{max} is the total length of the projection, and the double prime denotes the second derivative with respect to ξ. The two terms in this functional represent the competing goals of achieving a good fit to the data while maintaining a smooth curve, with the smoothing parameter α mediating the tradeoff. It can be shown that the minimizers of this functional are the curves known as natural cubic splines [12,13]. These are piecewise cubic curves that join at

the abscissa values ξ_i, where they are continuous up to and including the second derivative.

Clearly the choice of α determines the degree of smoothing that is applied to the data, and it is this parameter that is determined from the statistics of the data itself in an *adaptive* implementation of penalized least-squares smoothing using an algorithm known as generalized cross-validation [14]. Thus a generally different α is used in the smoothing of each of the M functions $P_k(\xi_i)$. The resulting continuous splines $P_k^s(\xi)$ must then be sampled to yield the discrete functions $P_k^s(\xi_i)$ which are subjected to the inverse DFT in step 3 above.

B. Ideal Observer Framework

The ideal-observer framework [15] offers a way of assessing the amount of information the data from an imaging device contains with regard to the performance of a specified task. For example, the simplest such task is the detection of a signal of known strength, shape, and location in a specified background. For linear imaging processes in which the noise in the output image is assumed to be additive, Gaussian, zero-mean, stationary, and independent of the presence or absence of the signal, the ideal-observer framework allows us to characterize the quality of the imaging system data with respect to the performance of the specified signal-detection task by a single number, the ideal-observer signal-to-noise ratio (SNR), which can be expressed as

$$SNR_I^2 = \int \frac{|\Delta S_{out}(\upsilon)|^2}{W(\upsilon)} d\upsilon, \quad (2)$$

where $|\Delta S_{out}(\upsilon)|^2$ is the power spectrum of the signal in output space, i.e. after it has been degraded by the imaging system, and $W(\upsilon)$ is the Wiener spectrum. By acquiring an ensemble of images containing the signal and the background as well as an ensemble of images containing the background alone, $|\Delta S_{out}(\upsilon)|^2$ can be easily determined from by computing the power spectrum of the difference between the two ensemble averages. Moreover, for a uniform background, the Wiener spectrum $W(\upsilon)$ can then be computed from the ensemble of background images.

C. Data Acquisition and Processing

Computing the ideal-observer SNR for a given imaging geometry and processing approach requires two ensembles of images: one set consisting of images of the signal and background together and one set consisting of images of the background alone. In order to preserve the flexibility to compute the ideal-observer SNR for lesions of varying concentration, we acquired high-count projection images of the signal alone (i.e., in a cold background), which could be scaled as desired and added to the ensemble of background projections alone to produce an ensemble of signal-plus-background projections. These were then either analyzed directly for the planar geometry or processed and reconstructed for the tomographic geometries.

Our phantom consisted of an 800 cc cylinder with a 1 cm outer diameter spherical lesion insert. This lesion size is representative of the smallest currently detected in scintimammography. For each geometry we acquired 20

images of the background alone, for which we put 3.7 mCi of activity in the phantom and imaged for 1 minute total for each conventional or dedicated SPECT acquisition and 30 seconds for each planar view. These combinations of activity and imaging time were chosen to produce clinically realistic count levels using the following reasoning. As with Wang *et al.*, we began with the assumption that 1% of a typical 25 mCi clinical dose of Tc-99m-Sestamibi is taken up by the myocardium. Using the volume of the myocardium in the Data Spectrum Corporation cardiac insert as a guide, along with the assumption that soft tissue will have a 1:15 concentration relative to the myocardium allowed us to determine the expected concentration of activity in healthy breast tissue. We wished to compare detectability in these three geometries given the same total imaging time, and assumed that typical clinical imaging times would be 30 minutes per SPECT study and 15 minutes per planar view. Thus we were comparing the three geometries for equal total imaging times given that two-view planar studies are common. We then scaled up the calculated concentration by a factor of 30 and scaled down the imaging times by the same factor to allow for more rapid data acquisition. All times were adjusted to compensate for the decay of the activity. To image the lesion we filled it with 7.6 mCi of Tc-99m, placed it in the cylinder filled with zero-activity water, and imaged for 30 minutes in conventional and dedicated SPECT and 20 minutes for the planar view. This combination of activity and imaging time was chosen to provide high-count, low-noise images of the signal, which could be scaled appropriately and added to the background images.

The dedicated SPECT images were acquired by placing the phantom at the center of rotation of a Picker XP2000 two-headed SPECT system with the heads rotating at the minimum radius of rotation (9.0 cm). In this configuration the heads were within 1.5 cm of the walls of the phantom. The breast phantom was not attached to an anthropomorphic torso phantom because Wang *et al.* showed that with proper shielding the contribution of scatter from the torso can be made negligible. We acquired 120 views over 360° with each head acquiring to a 128x128 matrix (pixel size=4.67 mm). We used a low-energy, ultra-high resolution collimator. The conventional SPECT images were also acquired in the absence of an anthropomorphic torso phantom, although the radius of rotation (25 cm) and the placement of the breast phantom (17 cm off-center) were determined with the torso phantom in place. The reason for this curious arrangement was to isolate the effect of the large radius of rotation on lesion detectability, without the additional degradations caused by attenuation or scatter in the torso. The ideal-observer SNR results for this arrangement thus represent an upper bound on the true detectability. This arrangement also facilitates computation of the Wiener spectrum, which requires images of a uniform background that would have been difficult to achieve in the presence of the highly non-uniform attenuation caused by the thorax. In other respects, the conventional SPECT images used the same acquisition parameters as the dedicated SPECT. Finally, the planar views were acquired with the phantom flush against one head, which acquired on a 128x128 matrix with a 2-fold magnification (pixel size=2.33 mm).

For the tomographic geometries, ideal-observer SNRs were computed from this data by the following procedure:

1. The signal projections were scaled to simulate a desired tumor-background concentration ratio (6:1 in this case) and added to each of the 20 sets of background projections.

2. The slice through the center of the lesion was selected and the 20 corresponding signal-plus-background sinograms were reconstructed by filtered backprojection using ramp and Hanning filters with various cutoff frequencies (0.4, 0.6, 0.8, and 1.0 times the Nyquist frequency). The sinograms were also processed using the effective 2D smoothing procedure described above and reconstructed by filtered backprojection. A zeroth-order Chang's method attenuation correction was applied to all reconstructed images.

3. The 20 corresponding sinograms of background alone were processed in the same ways.

4. An average signal image was determined by subtracting the average of the 20 background alone reconstructions from the average of the 20 signal plus background reconstructions. The signal power spectrum was computed by squaring the Fourier transform of this image.

5. The Wiener spectrum was computed from the 20 images of background alone by subtracting the average background image from each of the individual background images, resulting in 20 noise images. The uniform cylinders were isolated by multiplying each noise image by a circularly symmetric window of the form:

$$w(\mathbf{r}) = 1 \quad \text{for } |\mathbf{r}| \leq 0.9R,$$
$$w(\mathbf{r}) = 0.5 \times (1 + \cos(\pi(|\mathbf{r}| - 0.9R)/0.2R),$$
$$\text{for } 0.9R < |\mathbf{r}| < 1.1R, \text{ and} \tag{3}$$
$$w(\mathbf{r}) = 0 \quad \text{for } |\mathbf{r}| \geq 1.1R,$$

(with appropriate shifting for the off-center cylinder in the conventional geometry), where R is the radius of the cylinder and $\mathbf{r}$ the radial position in the image. The result was centered and zero-padded to 128x128. The noise power spectrum of each of the 20 images was computed by taking the square of the Fourier transform of the resulting image. The 20 noise power spectra were then averaged and scaled so that the volume under the Wiener spectrum equaled the average variance in the cylinders in the background images.

6. The ideal observer SNR was then determined by summing the quotient of the signal spectrum and Wiener spectrum.

For planar views, the procedure was essentially the same, without the reconstruction step but including the application of the effective 2D smoothing. Signal spectra were determined from the difference between the ensemble averages of signal-plus-background images and background images and Wiener spectra from the background images alone, using the same rolled-off cylindrical window to isolate a reasonably uniform region and minimize truncation effects.

Finally it should be noted that in using the ideal-observer framework at all it is implicitly being assumed that the data satisfy the assumptions discussed in section II.B: that the system is linear and that the noise in the planar or reconstructed images is additive, Gaussian, zero-mean, stationary, and independent of the presence or absence of the

signal. Given the reasonably high count levels (~10-20/pixel) the fact that the signal is relatively small and low contrast, and that fact that the background is reasonably uniform, the assumptions about the noise seem reasonable. The requirement of linearity seemingly undermines the use of the framework to analyze images that have been processed by adaptive, effective multi-dimensional smoothing. However, what is truly required for equation (2) to be meaningful is not linearity in the face of any possible input but more specifically that the system transfer function be the same whether the particular signal of interest is present or absent from the particular background of interest. Again, because the signal in question is relatively small and low contrast, it should not greatly affect the noise properties of the projection images and thus the effective multi-dimensional smoothing algorithm should yield a similar effective system transfer function whether or not the signal is present.

III. RESULTS

The ideal-observer SNRs for the detection of a 1-cm lesion with a 6:1 lesion-background concentration ratio are listed in Table 1 for the two tomographic geometries.

Table 1
Ideal Observer SNRs for conventional and dedicated SPECT

Processing Method	Dedicated Breast SPECT	Conventional SPECT
Hanning (cutoff=0.4)	8.0	3.0
Hanning (cutoff=0.6)	9.8	3.6
Hanning (cutoff=0.8)	10.3	3.7
Hanning (cutoff=1.0)	10.0	3.6
Ramp (cutoff=0.4)	9.5	3.7
Ramp (cutoff=0.6)	10.1	3.6
Ramp (cutoff=0.8)	9.7	3.4
Ramp (cutoff=1.0)	9.3	3.5
Eff. 2D smoothing	11.4	6.1

The ideal-observer SNR for the planar views was found to be 6.5 without processing and 6.7 after adaptive, effective 2D smoothing, again for a 1-cm lesion with a 6:1 lesion-background ratio. The threshold SNR necessary for a human to be able to detect reliably a signal in a noisy background is 5.0 [16]. Regardless of the processing approach, the conventional SPECT and planar geometries yield SNRs that are below or only just above this threshold. The results above thus confirm quantitatively the findings of Wang et al. that a 1-cm, 6:1 lesion is difficult or impossible to detect using planar or conventional SPECT geometries, but reliably detectable using a dedicated geometry. The results also indicate that effective 2D smoothing provides improvement in SNR over other filtering approaches. Images corresponding to the two different geometries for selected processing methods are shown in Figure 1. These confirm

visually the conclusions just stated: the 6:1 lesion is quite visible in all of the dedicated reconstructions, while it is effectively undetectable in the conventional reconstructions. The 6:1 lesion is rather difficult to discern in the unprocessed planar view depicted in Figure 2 and only slightly more visible in the smoothed view.
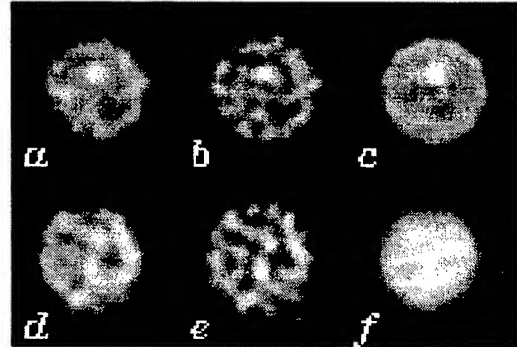

Figure 1: Reconstructed slices of a cylindrical phantom containing a 1-cm, 6:1 lesion for two different tomographic geometries. (a) Dedicated SPECT and Hanning filtering (cutoff=0.8). (b) Dedicated SPECT and ramp filtering (cutoff=0.6). (c) Dedicated SPECT and effective 2D smoothing. (d) Conventional SPECT and Hanning filtering (cutoff=0.8). (c) Conventional SPECT and ramp filtering (cutoff=0.6). (f) Conventional SPECT and effective 2D smoothing.
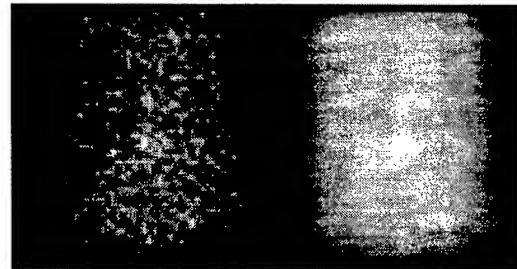

Figure 2: Planar images of a cylindrical phantom containing a 1-cm, 6:1 lesion. The image on the left is unprocessed; the image on the right has undergone adaptive, effective 2D smoothing. The lesion on the left is virtually undetectable, consistent with the low calculated SNR of 6.5. On the right, the lesion is slightly more visible, and corresponds to an SNR of 6.7.

IV. DISCUSSION AND CONCLUSIONS

The SNR values given in Table 1 suggest that a dedicated SPECT geometry would lead to improved detectability for clinically typical lesions over the planar and conventional SPECT geometries. In fact, in the presence of attenuation and scatter from the torso we would expect the difference between the dedicated and conventional SPECT geometries to be even greater than it is here. The success of the dedicated geometry can be attributed to the fact that it combines the advantages of the other two approaches: because of its small radius of rotation, it offers resolution comparable to that of a planar view acquired with the scintillation camera flush against the phantom, while it offers the improved contrast offered by a tomographic system's ability to separate lesion activity from overlying and underlying activity.

The SNR values also support the hypothesis that adaptive, effective multi-dimensional smoothing may

improve lesion detectability in tomographic scintimammography. It can be shown that any sort of linear processing performed on an image should not affect the ideal-observer SNR, because it modifies the signal and noise in equivalent and canceling ways. This sometimes surprising fact can be seen in the results of Table 1, for the SNRs for reconstructions using various Hanning and ramp filters show remarkably little variation despite the greatly different appearances of the resulting images. It is then reasonable to average these results together to give an SNR of 9.6 for dedicated SPECT using linear processing and 3.5 for conventional SPECT using linear processing. Because it can affect the signal and noise in different ways, non-linear processing can in principle alter the ideal observer SNR and we find that the adaptive, effective 2D smoothing leads to improvement, yielding SNRs of 11.4 for dedicated SPECT and 6.1 for conventional SPECT.

Finally, it should be noted that in the case of a linear imaging process, the calculated ideal-observer SNR should be directly proportional to the lesion concentration and thus to the tumor-background concentration ratio. This allows us to express the SNRs as a function of tumor-background concentration ratio, C, yielding an SNR of $1.6*C$ for the dedicated SPECT geometry using linear processing, an SNR of $0.6*C$ for the conventional SPECT geometry using linear processing, and an SNR of $1.1*C$ for the unprocessed planar geometry. Assuming that an SNR of 5.0 corresponds to the threshold of detectability, this allows us to conclude that for the specified imaging times and lesion size, the minimum tumor-background ratios needed for detectability are 3.1-to-1 for the dedicated geometry, 8.3-to-1 for the conventional geometry, and 4.5-to-1 for the planar geometry.

Throughout this study we have been comparing the three geometries on the assumption of equal total imaging times (30 minutes) and implicitly assuming in the planar geometry that we were examining the one of the two 15-minute views in which the lesion appeared most prominently. It is not necessarily possible to identify this view *a priori*, but assuming it were we can estimate the SNR corresponding to a 30-minute acquisition at that single view. Because SNRs for linear methods should in principle scale as the square root of the imaging time, the SNR would be approximately $6.5 * \sqrt{2} = 9.2$, or comparable to the dedicated SPECT SNR.

V. ACKNOWLEDGMENTS

VI. REFERENCES

[1] E. Bombardieri, F. Crippa, L. Maffioli and M. Greco, "Nuclear medicine techniques for the study of breast cancer," *Eur. J. Nucl. Med.,* vol. 24, pp. 809-824, 1997.

[2] A.D. Waxman, "The role of Tc-99m Methoxy-isobutylisonitrile in imaging breast cancer," *Seminars in Nuclear Medicine,* vol. 27, pp. 40-54, 1997.

[3] J. Maublant, M. de Latour, D. Mestas, A. Clemenson, S. Charrier, V. Feillel, G. Le Bouedec, P. Kaufmann, J. Dauplat and A. Veyre, "Technetium-99m-sestamibi uptake in breast tumor and associated lymph nodes," *J. Nucl. Med.,* vol. 37, pp. 922-925, 1996.

[4] L. Diggles, I. Mena and I. Khalkhali, "Technical aspects of prone dependent-breast scintimammography," *J. Nucl. Med. Tech.,* vol. 22, pp. 165-170, 1994.

[5] H. Palmedo, A. Schomburg, F. Grünwald, P. Mallman, D. Krebs and H.-J. Biersack, "Technetium-99m-MIBI scintimammography for suspicious breast lesions," *J. Nucl. Med.,* vol. 37, pp. 626-630, 1996.

[6] M.A. Nathan, J.E. Seabold, T. Barloon, J. Quesenberry, D.L. Bushnell, R.A. Robinson and D.C. Young, "Planar vs. SPECT Tc-99m MIBI evaluation of suspicious breast lesions on mammography: histologic correlation," *J. Nucl. Med.,* vol. 35, p. 229P, 1994.

[7] N. Nagaraj, A. Waxman, G. Ashok, S. Khan, L. Memsic, J. Yadegar and E. Phillips, "Comparison of SPECT and planar Tc-99m sestamibi (MIBI) imaging in patients with carcinoma of the breast," *J. Nucl. Med.,* vol. 35, p. 299P, 1994.

[8] R. Tiling, M. Pechmann, H. Sommer, R. Moser, K. Tatsch and K. Khan, "Does SPECT improve the diagnostic accuracy of planar scintimammography with sestamibi?' *J. Nucl. Med.,* vol. 23, pp. 97-102, 1982.

[9] H. Wang, C. Scarfone, K.L. Greer, R.E. Coleman and R.J. Jaszczak, "Prone breast tumor imaging using vertical axis-of-rotation SPECT systems: an intital study," *IEEE Trans. Nucl. Sci.,* vol. 44, pp. 1271-1276, 1997.

[10] A. Williams and K. Burrage, "Surface fitting using GCV smoothing splines on supercomputers," *CD-ROM proceedings of the 1995 ACM/IEEE conference on supercomputing,* 1995. The proceedings are also available on the World Wide Web at *http://www.supercomp.org/sc95/proceedings*.

[11] X. Pan, "A general approach for multidimensional smoothing," *Med. Phys.,* accepted pending minor revision (1997).

[12] P.J. Green and B.W. Silverman, *Nonparametric regression and generalized linear models,* London: Chapman Hall, pp. 11-27, 1994.

[13] J.A. Fessler, "Penalized weighted least-squares image reconstruction for positron emission tomography," *IEEE Trans. Med. Imag.,* vol. 13, pp. 290-300, 1994.

[14] D.M. Bates, M.J. Lindstrom, G. Wahba and B.S. Yandell, "GCVPACK routines for generalized cross validation," *Commun. Statist. Simul. Comput.,* vol. 16, pp. 263-297, 1987.

[15] "Medical imaging—the assessment of image quality," *ICRU Report 54,* 1996. Available from International Commission on Radiation Units and Measurements, 7910 Woodmont Ave., Bethesda, MD, 20814.

[16] A. Rose, *Vision: Human and Electronic,* New York: Plenum Press, 1973.

Direct Spline-Based Inversion of the Three-Dimensional Radon Transform with Application to Cardiac Phantom Data

P.J. La Rivière and X. Pan

Department of Radiology, The University of Chicago, Chicago, IL, 60637

Abstract

While the exact inverse three-dimensional Radon transform is a continuous integral equation, the discrete nature of the data output by tomographic imaging systems demands that images be reconstructed using a discrete approximation to the transform. However, by fitting an analytic function to the planar-integral data prior to reconstruction, one can avoid such approximations and preserve and exploit the continuous nature of the inverse transform.

We present methods for the evaluation of the inverse 3D Radon transform in which cubic spline functions are fit to the planar-integral data, allowing exact computation of the second derivative that appears in the inversion formula and also eliminating the need for interpolation upon backprojection. This approach is theoretically intriguing and has the advantage of directness when one wishes to smooth noisy data prior to reconstruction. In this case, a *smoothing* spline can be fit to the data and reconstruction can proceed directly from the spline coefficients. We find that the 3D direct-spline algorithm has superior resolution to 3D filtered backprojection, albeit with higher noise, and that it has a slightly lower ideal-observer signal-to-noise ratio for the detection of a 1-cm, spherical lesion with a 6:1 lesion-background concentration ratio.

I. INTRODUCTION

The inverse Radon transform in two or three dimensions provides the mathematical foundation for tomographic imaging, which involves reconstructing images of distributions of anatomical or physiological properties from projections of those distributions. We will denote any such distribution in two dimensions by $f(x,y)$ and label each projection through it by the pair $\{\xi, \varphi\}$, where φ specifies the *projection angle* and ξ the *projection distance*. The value of such a projection is given by the line integral of the distribution along the line specified by $\{\xi, \varphi\}$ and will be denoted by $p(\xi, \varphi)$. Similarly, we will denote a 3D distribution by $f(x,y,z)$ and label each projection through it by the triplet $\{\xi, \theta, \varphi\}$, where θ and φ specify the orientation of the plane and ξ the distance of the plane to the origin of the coordinates. The value of such a projection is given by the *planar* integral of the distribution over the plane specified by $\{\xi, \theta, \varphi\}$ and will be denoted by $p(\xi, \theta, \varphi)$.

The Radon transform is a continuous, integral transform that relates the sinogram $p(\xi, \varphi)$ to $f(x,y)$ in two dimensions and the sinogram $p(\xi, \theta, \varphi)$ to $f(x,y,z)$ in three dimensions [1]. Inverting the Radon transform exactly to recover a distribution requires continuous, noise-free knowledge of the distribution's sinogram, which entails having an infinite set of perfect projection measurements. In practice, of course, one can only collect projection data in the 2D case for a finite number of projection distances ξ_i (we call these discrete samples *projection bins*) at a finite number of projection angles φ_j, and the measurements are invariably contaminated with noise. In the 3D case, the planar-integral projection data cannot generally be measured directly and must instead be generated from line-integral projection data; it is, however, still only generated for a finite number of projection bins ξ_i and projection angles φ_j and θ_k.

Because the sinogram $p(\xi, \varphi)$ in two dimensions (or $p(\xi, \theta, \varphi)$ in three dimensions) is known only on a finite grid of ξ_i and φ_j (or ξ_i, φ_j, and θ_k), we cannot invert the Radon transform exactly to recover the distribution $f(x,y)$ (or $f(x,y,z)$); we must instead turn to a discrete approximation of the inverse. For instance, one way of implementing the most popular 2D Radon inversion algorithm—filtered backprojection (FBP) [2]—begins with a discrete filtration of the sinogram, accomplished by taking a discrete Fourier transform of the sinogram with respect to projection bin ξ_i and multiplying by a discrete ramp filter, perhaps apodized by a window function to control noise. After taking an inverse Fourier transform to return to $\{\xi, \varphi\}$ space, one backprojects the filtered samples of ξ for each projection angle φ_j onto the image grid and sums the resulting images to obtain a final reconstructed image. One view of the process of backprojection is from the standpoint of a pixel-by-pixel traversal of image space, in which a perpendicular ray is cast from each pixel to each projection angle in turn, at which point an appropriate sinogram value is picked up. In this view, the difficulty lies in determining exactly what value to pick up from each projection angle, for in general the perpendicular line will not fall directly in the center of a projection bin. In the simplest schemes, one simply picks up the value of the bin the pixel projects onto, while in a more complicated approach one might perform a linear interpolation of the values in the two nearest projection bins. In the most sophisticated schemes, one uses a weighted average of the values in a slightly larger neighborhood. There are of course other ways of viewing the backprojection step in FBP [3], but they all must contend with the fact that we only have a finite number of projection bins at each projection angle. A similar procedure can be used to implement 3D FBP [4].

The discreteness of the sinogram thus dictates discreteness in both the filtration and backprojection steps of the algorithm. If, however, one had a continuous, analytic expression for the sinogram at each projection angle—if the sinogram were a set of known one-dimensional functions of ξ—it might be possible to implement the filtration and backprojection steps in a continuous manner. The filtration could be performed analytically, and the resulting filtered projections would be continuous functions which, in the

pixel-traversal view of backprojection described above, could be sampled wherever a projection might strike without any need to interpolate. Naturally, such an analytic, continuous expression for the sinogram cannot be obtained directly from any tomographic imaging system, but must rather be obtained by fitting an analytic expression to the discretely sampled data. Not every class of function that could be fit to the data would allow the filtration of the projections to be calculated in closed form, but Wahba has shown that one class that does behave nicely in the 2D case are the cubic splines, piecewise third-order polynomials that are continuous up to and including the second derivative at the connection points between pieces [5]. This is the class of fitting functions we investigate in this paper, introducing Wahba's results, and extending the treatment to the 3D Radon transform.

This method offers a certain conceptual appeal, as well as the advantage of directness in one situation of practical concern—reconstruction from data that has been smoothed to reduce noise. When dealing with noisy data, the functions being fit to the data need not be strictly interpolating functions but could rather be smoothing functions. Conveniently, the natural cubic splines, mentioned above as permitting an analytic inversion of the Radon transform, are also the class of functions that minimize the penalized least-squares measure frequently used in smoothing noisy data [6]. While it is of course possible to smooth noisy data with cubic splines and then sample a discrete sinogram for use in a discrete reconstruction algorithm, the proposed method allows for a direct reconstruction from the coefficients of the smoothing spline.

II. METHODS

A. Inverse 2D Radon transform in coordinate space

The essential problem in 2D tomography is the reconstruction of a distribution $f(x,y)$ (or $f(r,\theta)$ in polar coordinates) from knowledge of the discrete sinogram $p(\xi_i,\varphi_j)$, where $i=-N,...,N$ and $j=1,...,M$. This convention for the index i, particularly the choice of an odd number of projection bins, will simplify the mathematical expressions to be derived below, but the proposed technique is applicable to geometries with an even number of bins as well. We assume in the present paper that we have a parallel-beam geometry.

The inverse 2D Radon transform may be expressed in coordinate space as

$$f(r,\theta)=\frac{1}{2\pi^2}\int_0^\pi J_{r,\theta}(\varphi)\,d\varphi, \qquad (1)$$

where

$$J_{r,\theta}(\varphi)=\left\{\lim_{\varepsilon\to 0}\left[\int_{-\infty}^{\xi'-\varepsilon}\frac{p'(\xi,\varphi)}{\xi'-\xi}\,d\xi+\int_{\xi'+\varepsilon}^{\infty}\frac{p'(\xi,\varphi)}{\xi'-\xi}\,d\xi\right]\right\}, \qquad (2)$$

and in which $\xi'=r\cos(\theta-\varphi)$ and $p'(\xi,\varphi)$ is the first derivative of the sinogram $p(\xi,\varphi)$ with respect to ξ [7]. Taking the limit as $\varepsilon\to 0$ of the sum of these two integrals allows us to avoid integrating over the singularity at $\xi'=\xi$.

In general equations (1) and (2) are less useful than other expressions for the 2D inverse Radon transform because the data collected in PET, SPECT, or CT constitute samples of the sinogram $p(\xi,\varphi)$ itself, and do not provide any direct information about $p'(\xi,\varphi)$. However, by fitting an analytic, differentiable function of ξ to the projection data at each angle, we could obtain an expression for $p'(\xi,\varphi)$. If $p'(\xi,\varphi)$ had an auspicious functional form, we would then be able to solve the integrals in equation (2) in closed form. Wahba has shown that the natural cubic splines to be discussed below allow such a solution and she derived a complicated closed-form expression for the inverse in terms of the coefficients of the spline, in which care must be taken in handling potentially unstable terms, and which leads to a very computationally intense image reconstruction algorithm [5]. In contrast, we shall see that a direct-spline inverse of the 3D Radon transform has a very simple form, has no instabilities, and is computationally efficient.

B. Inverse 3D Radon transform in coordinate space

The essential problem in 3D computed tomography is the reconstruction of a distribution $f(x,y,z)$ from knowledge of the discrete planar-integral sinogram $p(\xi_i,\theta_k,\varphi_j)$, where $i=-N,...,N,\cdot\;\;j=1,...,M_\varphi$, and $k=1,...,M_\theta$ [4]. In general, these planar integrals are not measured directly by tomographic imaging systems, but must rather be calculated by "rebinning" the line integrals that are measured directly [4].

The inverse 3D Radon transform has a form similar to the 2D case, with a few differences that greatly simplify the task of evaluating it numerically. Specifically,

$$f(x,y,z)=\frac{-1}{4\pi^2}\int_0^\pi\int_0^\pi p''(\xi',\theta,\varphi)\,\sin\theta\,d\varphi\,d\theta, \qquad (3)$$

where

$$\xi'=x\sin\theta\cos\varphi+y\sin\theta\sin\varphi+z\cos\theta, \qquad (4)$$

and $p''(\xi',\theta,\varphi)$ is the second derivative of the 3D sinogram with respect to ξ'. This expression differs in two principal ways from the expression for the 2D inverse Radon transform given by equations (1) and (2). First, it now involves the second derivative of the sinogram with respect to ξ' rather than the first derivative and second, the convolution in ξ has disappeared. This reflects the fact that an inverse Radon transform of odd degree can be calculated using purely *local* information—the value of the image at a point (x,y,z) can be determined solely from information at points in the sinogram space that (x,y,z) projects onto, rather than from a convolution integral over all points in sinogram space as in the even-dimensional case [1].

C. Spline-based inverse of the 3D Radon transform

To obtain an expression for the planar-integral sinogram that is analytic and differentiable with respect to the variable ξ, we fit a function $\hat{p}_{\varphi_j,\theta_k}(\xi)$ to the planar-integral sinogram values $p_{\varphi_j,\theta_k}(\xi_i)$ measured on the $2N+1$ abscissas ξ_i $(i=-N,...,N)$ for each pair of angles $\{\varphi_j,\theta_k\}$. If the data is noiseless, it is desirable to use a function that passes through the points $p_{\varphi_j,\theta_k}(\xi_i)$, which we call an *interpolating* curve, while if the data is noisy a *smoothing* curve may be

more appropriate. One fitting framework that can handle both of these situations is known as penalized least-squares [6], in which the function $\hat{p}_{\varphi_j,\theta_k}(\xi)$ (written as $\hat{p}(\xi)$ for compactness) is chosen to be the minimizer of the functional

$$F[\hat{p}(\xi)] = \sum_{i=-N}^{N}[p(\xi_i) - \hat{p}(\xi_i)]^2 + \lambda \int_{a}^{b}(\hat{p}''(\xi))^2 d\xi, \quad (5)$$

where a and b are the endpoints of the interval on which the curve $\hat{p}(\xi)$ is to be defined. The parameter λ is known as the smoothing parameter and can either be chosen *a priori* [8] or determined automatically from the statistics of the data using a procedure such as generalized cross-validation (GCV) [9].

The minimizers of this functional F are functions known as natural cubic splines [8]. These are piecewise polynomial curves that join at the abscissa values ξ_i, where they are continuous up to and including the second derivative. They can be represented as

$$\hat{p}_{\varphi_j,\theta_k}(\xi) = a_i + b_i\xi + c_i\xi^2/2 + d_i\xi^3/3, \quad \xi\,\varepsilon\,[\xi_i,\xi_{i+1}], \quad (6)$$

where a_i, b_i, c_i, and d_i are constants that fully specify the spline on the interval $[\xi_i,\xi_{i+1}]$. Of interest to the Radon inversion problem is the fact that we can approximate the second derivative of the sinogram for fixed angles $\{\varphi_j,\theta_k\}$ by

$$p''(\xi,\theta_k,\varphi_j) \cong \hat{p}''_{\varphi_j,\theta_k}(\xi) = c_i + 2d_i\xi, \quad \xi\,\varepsilon\,[\xi_i,\xi_{i+1}]. \quad (7)$$

We can substitute this expression into equation (3), expressing the integrals as sums to reflect the fact that the sinogram is still discrete in the angular variables. This yields

$$f(x,y,z) = -\frac{1}{4}\frac{1}{M_\theta}\frac{1}{M_\varphi}\sum_{j=1}^{M_\varphi}\sum_{k=1}^{M_\theta}J_{x,y,z}(\theta_k,\varphi_j)\sin\theta_k, \quad (8)$$

where

$$J_{x,y,z}(\theta_k,\varphi_j) = c_i + 2d_i\xi', \quad \xi'\,\varepsilon\,[\xi_i,\xi_{i+1}], \quad (9)$$

and

$$\xi' = x\sin\theta_k\cos\varphi_j + y\sin\theta_k\sin\varphi_j + z\cos\theta_k. \quad (10)$$

The simplicity of the 3D inversion is now apparent, for the functions J can be evaluated in a straightforward manner for a given point (x,y,z) in image space.

D. Application to phantom and real data

To test the 3D direct-spline inversion algorithm, we reconstructed images of a Data Spectrum ventricular phantom from projection data acquired on a Picker XP3000 three-headed SPECT system fitted with low-energy, high-resolution, parallel-hole collimators. The phantom was filled with 3.27 mCi of Tc-99m and placed at the center of rotation. Each head collected data on a 128x128 grid and at 120 projection angles over 360°. We rebinned the projection data from a single head to generate planar integrals on a 128x60x120 grid. Images were reconstructed from this planar-integral data using 3D FBP, and also using the 3D direct-spline inversion method. We then fit smoothing splines to the planar-integral data and reconstructed directly from the spline coefficients. Finally, we sampled the smoothing spline to generate a discrete sinogram and used that as input to the 3D FBP algorithm.

E. Resolution, noise, and signal-to-noise studies

In order to compare quantitatively the resolution of the 3D direct-spline algorithm with 3D FBP, we acquired projection images of a small (1 cm) spherical lesion containing 7.6 mCi of Tc-99m placed in an 800 cc cylindrical phantom containing cold (zero-activity) water. A Picker XP2000 two-headed SPECT system fitted with ultra-high-resolution, parallel-hole collimators was used. The heads rotated at their minimum radius of rotation (9 cm) and acquired 120 views over 360° onto a 128x128 matrix (pixel size=4.67 mm).

We rebinned the projection data to generate planar-integral data on a 128x60x120 grid and then reconstructed the slice through the center of the lesion using 3D FBP and the 3D direct-spline method (using interpolating splines). The reconstructed lesion was approximately a symmetric 2D Gaussian in shape and we determined its full-width half-maximum by collapsing it into a one-dimensional function and fitting this profile with a Gaussian curve.

In order to isolate the contribution of the reconstruction algorithm to the FWHM of the lesion in the reconstructed images, the contribution from the lesion's inherent width as well as the imaging system's point-spread function had to be removed. The net effect of these two factors was estimated by determining the average FWHM of the lesion as it appeared in the 120 projection images. Assuming then that the reconstruction algorithms can be characterized by Gaussian point-spread functions, the FWHM of these functions was determined by subtracting (in quadrature) the average projection FWHM from the FWHMs of the reconstructed lesions discussed above.

To characterize the noise level in images reconstructed by the direct-spline method and FBP, we acquired 20 1-minute projection datasets of the same 800 cc cylindrical phantom used in the previous section, this time containing 3.7 mCi of Tc-99m and no lesion. Images of one slice of this uniform cylinder were reconstructed using 3D FBP and 3D spline inversion using interpolating splines. For a given algorithm, six circular regions of interest (ROI) were examined in each of the 20 slices and the coefficient of variation (the standard deviation of the pixel values in the ROI divided by the mean of the pixel values in the ROI) calculated for each of the 120 ROIs. The average of these 120 coefficients of variation was then computed.

It is not uncommon for a reconstruction algorithm to offer enhanced resolution at the price of amplified noise. The overall effect of such a tradeoff is sometimes better characterized by computing a signal-to-noise ratio (SNR). We used the two datasets described above to compute so-called ideal-observer SNRs. The ideal-observer framework [10] offers a way of assessing the amount of information the data output by an imaging device contains with regard to the performance of a specified task. For example, the simplest such task is the detection of a signal of known strength, shape, and location in a specified background. For linear imaging processes in which the noise in the output image is assumed to be additive, Gaussian, zero-mean, stationary, and independent of the presence or absence of the signal, the ideal-observer framework allows us to characterize fully the

quality of the imaging system data with respect to the performance of the specified signal-detection task with a single number, the ideal observer signal-to-noise ratio (SNR), which can be expressed as

$$SNR_I^2 = \int \frac{|\Delta S_{out}(\upsilon)|^2}{W(\upsilon)} d\upsilon, \qquad (11)$$

where $|\Delta S_{out}(\upsilon)|^2$ is the power spectrum of the signal in output space, i.e. after it has been degraded by the imaging system, and $W(\upsilon)$ is the Wiener spectrum. We can view the 20 sets of projections of the cylinder alone used in the noise study as an ensemble of so-called background images, and by adding suitably scaled projections of the lesion used in the resolution study (the scaling was chosen to produce a 6:1 lesion-background ratio) we generated an ensemble of signal-plus-background images. Then $|\Delta S_{out}(\upsilon)|^2$ could be easily determined by computing the power spectrum of the difference between the two ensemble averages. The Wiener spectrum $W(\upsilon)$ could then be computed from the ensemble of background images.

The detailed procedure for the calculation of the ideal-observer SNR was as follows:

1. The lesion projections were scaled to simulate a desired lesion-background concentration ratio (6:1 in this case) and added to each of the 20 sets of background projections.

2. Images of the slice through the center of the lesion were reconstructed for the 20 signal-plus-background datasets using four different methods: the 3D spline-based inverse using interpolating splines, 3D FBP, the 3D spline-based inverse using smoothing splines, and 3D FBP using a sinogram sampled from these smoothing splines.

3. The 20 corresponding sinograms of background alone were reconstructed in the same four ways.

4. An average signal image was determined for each algorithm by subtracting the average of the 20 background reconstructions from the average of the 20 signal-plus-background reconstructions. The signal power spectrum was computed by squaring the Fourier transform of this image.

5. The Wiener spectrum for each reconstruction method was computed from the 20 background images by subtracting the average background image from each of the individual background images, resulting in 20 noise images. The uniform cylinders were isolated by multiplying each image by a circularly symmetric window with a cosine rolloff. The noise power spectrum of each of the 20 images was computed by taking the square of the Fourier transform of the resulting image. The 20 noise power spectra were averaged and scaled so that the volume under the resulting Wiener spectrum equaled the average variance in the cylinders.

6. The ideal observer SNR was then determined by summing the quotient of the signal spectrum and Wiener spectrum.

Finally, it should be noted that in using the ideal-observer framework at all it is implicitly being assumed that the data satisfies the assumptions discussed above: that the system is linear and that the noise in the planar or reconstructed images is additive, Gaussian, stationary, zero-mean, and independent of the presence or absence of the signal. Given the reasonably high count levels (~10-20/pixel) and the fact that the signal is relatively small and low contrast, these assumption about the noise are not unreasonable. The requirement of linearity seemingly undermines the use of the framework to analyze images that are reconstructed from smoothing splines that have been fit using an adaptive, and thus non-linear algorithm. However, what is truly required for equation (11) to be meaningful is not linearity in the face of any possible input but more specifically that the system transfer function be the same whether the particular signal of interest is present or absent from the particular background of interest. Again, because the signal in question is relatively small and low contrast, it should not greatly affect the noise properties of the projection images and thus the use of smoothing splines should yield a similar effective system transfer function whether the signal is present or absent.

III. RESULTS

The results of reconstructing the ventricular phantom data using 3D direct-spline inversion with interpolating splines, 3D FBP, 3D direct-spline inversion with smoothing splines, and 3D FBP using a sinogram resampled from the smoothing splines are depicted in Figure 1. The algorithms are seen to yield qualitatively similar results.

The calculated in-plane FWHM of the direct-spline inversion algorithm was found to be 3.9 mm compared to 5.0 mm for 3D FBP.

The coefficient of variability for the direct-spline inversion algorithm was found to be 0.34 compared to 0.23 for 3D FBP.

Finally, the ideal-observer signal-to-noise ratio results are summarized in Table 1 for the two basic algorithms as well as their counterparts using smoothing splines. We observe that the direct-spline algorithm has a slightly lower SNR than FBP when using interpolating splines, and a slightly higher SNR when using smoothing splines. Moreover, the SNRs for both approaches worsen when using smoothing splines. Typical images reconstructed using each of these four methods are shown in Figure 2.

Table 1
Ideal-observer SNRs for various reconstruction algorithms

Algorithm	Ideal-observer SNR
3D interpolating spline	9.2
3D FBP	9.7
3D smoothing spline	7.6
3D FBP w/ smth. spline	7.2

IV. DISCUSSION AND CONCLUSIONS

As discussed in section I, the principal difference between the direct-spline inversion algorithm and FBP is in the nature of the interpolation step upon backprojection. In our implementation of FBP, the interpolation is linear, while in direct-spline inversion it is implicitly a more sophisticated cubic-spline interpolation. The cruder linear interpolation is

more likely to smooth over high-frequency variations in the projection data than is cubic-spline interpolation, and thus it is not surprising that the FBP algorithm has inferior resolution to the spline-based algorithm. However, because the high-frequency components of the data include considerable noise as well, the FBP algorithms would be expected to produce less noisy reconstructions than the spline-based reconstructions. This expectation is confirmed by the results of the noise study.
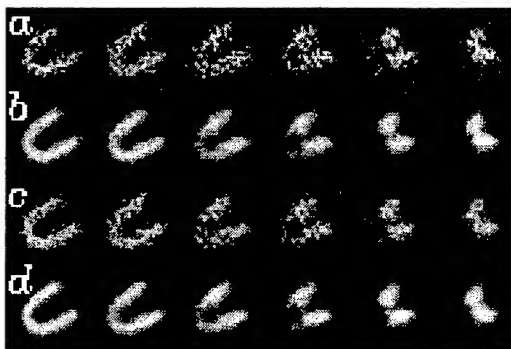

Figure 1. Selected slices of a ventricular phantom reconstructed by (a) 3D direct-spline inversion using interpolating splines, (b) 3D direct-spline inversion using smoothing splines, (c) 3D FBP, and (d) 3D FBP from a sinogram obtained by sampling the smoothing splines in (b).
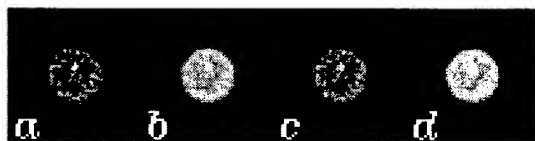
Figure 2. Reconstructions of a selected slice of a cylindrical phantom containing a spherical lesion. Reconstruction methods: (a) 3D direct-spline inversion using interpolating splines, (b) 3D direct-spline inversion using smoothing splines, (c) 3D FBP, and (d) 3D FBP from a sinogram obtained by sampling the smoothing splines in (b).

The results of Table 1 indicate that ideal-observer SNRs are slightly lower for the 3D direct-spline inversion using interpolating splines than for FBP. While strictly speaking ideal-observer SNRs should not be affected by linear processing, given the approximations we have made these results at least suggest that for this particular detection task, the spline algorithm's amplification of noise outweighs the improvement it affords in resolution relative to FBP. However, when the noise is mitigated prior to reconstruction, as when the projection data is fitted with smoothing splines, the SNR gap between the spline algorithm and FBP is reversed. This suggests that the spline-based algorithms may be of greatest use when resolution is paramount and the data contains relatively little noise, a situation more often encountered in CT than in nuclear medicine.

The use of smoothing splines is seen to worsen the SNR for both reconstruction methods. Indeed, it is clear from examining the images of Figure 2 that the reconstructed images using smoothing splines have an oversmoothed appearance. This can most likely be attributed to the fact that the 3D reconstruction process effectively involves a prior

smoothing during the rebinning step, when area-weighted forward projection is used to generate the planar-integral data from the projection data at each angle. The adaptive smoothing algorithm certainly makes allowances for the lower variability in the rebinned data and smooths this data less than it would the raw projection data. However, the modified statistics of the rebinned data simply do not agree as well with the statistical model assumed by the smoothing algorithm, which might thus be expected to yield a sub-optimal result. It remains a topic for further investigation as to whether smoothing the raw projection data prior to the rebinning step produces a better result.

V. ACKNOWLEDGMENTS

VI. REFERENCES

[1] S. Deans, The Radon Transform and Some of its Applications, New York: Wiley, 1983.

[2] L.A. Shepp and B.F. Logan, "The Fourier reconstruction of a head section," IEEE Trans. Nucl. Sci., vol. 21, pp. 21-43, 1974.

[3] R.H. Huesman, G.T. Gullberg, W.L. Greenberg and T.F. Budinger, User Manual, Donner Algorithms for Reconstruction Tomography, Lawrence Berkeley Laboratory, University of California, 1977.

[4] C. Wu, "Fully three-dimensional reconstruction in PET and SPECT by the use of three-dimensional Radon transforms," PhD Thesis, University of Chicago, 1994.

[5] G. Wahba, "A new approach to the numerical evaluation of the inverse Radon transform with discrete, noisy data," Technion Preprint Series MT-479, 1980.

[6] P.J. Green and B.W. Silverman, Nonparametric Regression and Generalized Linear Models, London: Chapman Hall, 1994.

[7] G.T. Herman and A. Naparstek, "Fast image reconstruction based on a Radon inversion formula appropriate for rapidly collected data," SIAM J. Applied Mathematics, vol. 33, pp. 511-533, 1977.

[8] J.A. Fessler, "Tomographic reconstruction using information-weighted spline smoothing," in Information Processing in Medical Imaging, H.H. Barrett and A.F. Gmitro, eds., Berlin: Springer-Verlag, 1993, pp. 290-310.

[9] D.M. Bates, M.J. Lindstrom, G. Wahba and B. Yandell "GCVPACK routines for generalized cross-validation," Commun. Statist. Simul. Comput., vol. 16, pp. 263-297, 1987.

[10] "Medical imaging—the assessment of image quality," ICRU Report 54, 1996. Available from ICRU, 7910 Woodmont Ave., Bethesda, MD, 20814.

Mathematical equivalence of zero-padding interpolation and circular sampling theorem interpolation with implications for direct Fourier image reconstruction

P. J. La Rivière and X. Pan

Department of Radiology, The University of Chicago, Chicago, IL, 60637

ABSTRACT

The speed and accuracy of Direct Fourier image reconstruction methods have long been hampered by the need to interpolate between the polar grid of Fourier data that is obtained from the measured projection data and the Cartesian grid of Fourier data that is needed to recover an image using the 2D FFT. Fast but crude interpolation schemes such as bilinear interpolation often lead to unacceptable image artifacts, while more sophisticated but computationally intense techniques such as circular sampling theorem (CST) interpolation negate the speed advantages afforded by the use of the 2D FFT. One technique that has been found to yield high-quality images without much computational penalty is a hybrid one in which zero-padding interpolation is first used to increase the density of samples on the polar grid after which bilinear interpolation onto the Cartesian grid is performed. In this work, we attempt to account for the success of the approach relative to the CST approach in three ways. First and most importantly, we establish that zero-padding interpolation of periodic functions that are sampled in accordance with the Nyquist criterion—precisely the sort of function encountered in the angular dimension of the polar grid—is exact and equivalent to circular sampling theorem interpolation. Second, we point out that both approaches make comparable approximations in interpolating in the radial direction. Finally, we indicate that the error introduced by the bilinear interpolation step in the zero-padding approach can be minimized by choosing sufficiently large zero-padding factors.

Keywords: Interpolation, Image Reconstruction, Direct Fourier Methods, Zero-padding, Circular Sampling Theorem, FFT

1. INTRODUCTION

The direct Fourier approach to tomographic image reconstruction has long had the potential to be among the most computationally efficient of image reconstruction algorithms because it harnesses the speed of the inverse 2D Fast Fourier Transform (FFT). The inverse 2D FFT is used to recover an image from samples of its Fourier transform that are obtained by Fourier transforming the projections of the object slice being imaged. As is widely known, the chief impediment to effective direct Fourier algorithms is the fact that the transformed projection data yields samples of the image transform that lie on a polar grid, while the inverse 2D FFT algorithm requires samples to lie on a Cartesian grid. The necessary interpolation between the two grids generally compromises either image quality, when a fast but crude interpolation is performed, or processing time, when a more sophisticated but computationally intense interpolation is performed. For instance, the simplest of interpolation schemes—nearest neighbor or bilinear interpolation—can lead to artifacts in the reconstructed images [1], while more accurate techniques, such as the gridding approaches of O'Sullivan [2] and Jackson et al. [3], involve more expensive convolutions over larger neighborhoods.

Stark et al. [4,5] have examined an interpolation scheme that is essentially exact when the polar samples of the transform satisfy certain conditions. They use truncated Whittaker-Shannon sinc interpolation in the radial direction and circular sampling theorem (CST) interpolation in the angular direction. CST interpolation is an exact interpolation method that follows from Whittaker-Shannon interpolation when the function being interpolated is periodic and sampled in accordance with the Nyquist criterion. Unfortunately, CST interpolation also involves a very expensive convolution, a processing burden Stark et al. attempted to mitigate by truncating the CST series, with a consequent tradeoff in accuracy.

At least one interpolation approach has been developed that generates images of comparable quality to CST and other convolution-based techniques with none of their computational drawbacks. It is a hybrid method in which zero-padding interpolation is first used to increase the density of polar samples in both the radial and angular directions, after which bilinear interpolation onto the Cartesian grid is performed. Zero-padding interpolation works by extending the discrete Fourier transform (DFT) of a finite sequence with zeroes and then performing an inverse DFT, yielding a sequence with additional, interpolated samples between the measured values of the original sequence. Because it exploits the FFT algorithm, zero-padding interpolation is very computationally efficient. The approach was studied explicitly and compared to CST-based techniques by Kak and Pan in the context of ultrasound diffraction tomography, where a similar Fourier-space interpolation problem arises [6].

In this work, we account for the surprising success of this hybrid approach in three ways. First, we establish that for the sort of periodic, bandlimited functions encountered in the angular dimension of the polar grid, zero-padding interpolation is equivalent to CST interpolation insofar as the values zero-padding interpolation generates on the denser polar grid it produces match the values CST interpolation would produce at those points. Of course, CST interpolation is not constrained to interpolate onto a polar grid of increased angular density—it can interpolate at arbitrary angular points—and thus avoids the additional linear interpolation needed in the zero-padding approach. However, we point out that the error arising from this last interpolation can be made negligibly small if the polar grid is made sufficiently dense through zero-padding. Finally, we indicate that both the zero-padding and CST approaches make comparable approximations in interpolating in the radial direction of the polar grid.

We begin in part 2 by reviewing the basic theory of CST and zero-padding interpolation and then proving their equivalence for one-dimensional, periodic functions sampled in accordance with the Nyquist criterion. In part 3, we introduce the specific interpolation problem encountered in direct Fourier image reconstruction and discuss the solution offered by CST and zero-padding interpolation, as well as simple bilinear interpolation. In part 4, we present the results of using these three methods to reconstruct images of numerical phantoms, both ideal and contaminated by Poisson noise.

2. THEORY

In this section we focus on the interpolation of one-dimensional, bandlimited, periodic functions of period 2π, introducing the CST and zero-padding approaches and then establishing their agreement on the grid of zero-padding interpolated values. The decision to focus on functions of period 2π simplifies notation and reflects the fact that this is the period of the angular samples encountered in direct Fourier reconstruction, but all of the results derived apply equally well to functions with other periods.

2.1. Circular Sampling Theorem

Circular Sampling Theorem interpolation is a special case of exact Whittaker-Shannon interpolation that applies to periodic functions sampled in accordance with the Nyquist criterion [7,8]. Consider a function $x(\theta)$ that is periodic with period 2π and bandlimited to frequency K (i.e., the coefficients of the function's Fourier series expansion satisfy $a_k = 0$ for $|k| > K$). If there are N equidistant samples of $x(\theta)$, i.e., $x(2\pi n/N)$, $n = 0,...,N-1$, and N is odd and $\geq 2K+1$, the circular sampling theorem (CST) states that the value of $x(\theta)$ may be determined exactly at arbitrary θ using

$$x(\theta) = \sum_{n=0}^{N-1} x\left(\frac{2\pi n}{N}\right) \frac{\sin\left[\frac{N}{2}\left(\theta - \frac{2\pi n}{N}\right)\right]}{N \sin\left[\frac{1}{2}\left(\theta - \frac{2\pi n}{N}\right)\right]}. \tag{1}$$

Similarly, if N is even and $\geq 2K$, the value of $x(\theta)$ may be determined exactly at arbitrary θ using

$$x(\theta) = \sum_{n=0}^{N-1} x\left(\frac{2\pi n}{N}\right) \frac{\sin\left[\frac{1}{2}(N-1)\left(\theta - \frac{2\pi n}{N}\right)\right]}{N \sin\left[\frac{1}{2}\left(\theta - \frac{2\pi n}{N}\right)\right]}. \tag{2}$$

2.2. Zero-padding interpolation

It is well known that extending the discrete Fourier Transform (DFT) of a finite sequence with zeroes and performing an inverse DFT increases the density of samples in the conjugate domain, thereby yielding interpolated values at regular intervals between the sequence's measured points. Fraser has given a thorough and enlightening discussion of the process that highlights a few of its more subtle points [9]. It is not widely known, however, that interpolation by zero-padding the DFT is equivalent to CST interpolation (and thus to exact Whittaker-Shannon interpolation) when the sequence in question represents samples of a periodic, bandlimited function sampled in accordance with the Nyquist criterion. To establish this equivalence, we begin with a rigorous definition of the zero-padding process, taking care to distinguish between the cases when the number of samples N is even or odd. Following Fraser, we define a superscript ⁻, such that $N^- = N$ (N even) and $N^- = N-1$ (N odd). Assume again that we have a function $x(\theta)$ that is periodic with period 2π and bandlimited to frequency K, and of which we have N equidistant samples, $x(2\pi n/N)$, $n = 0,...,N-1$, where $N \geq 2K$. The DFT of this sequence is given by ·

$$c_k = \frac{1}{N} \sum_{n=0}^{N-1} x\left(\frac{2\pi n}{N}\right) \exp(-j2\pi nk / N), \quad k = 0,...,N-1, \tag{3}$$

where $j = \sqrt{-1}$. Zero-padding involves the creation of a new sequence $d_{k'}$, having $L = P \cdot N$ terms, where P is an integer. Taking the inverse DFT of the new sequence yields a more densely sampled version of the original sequence. We need only explicitly build the first half of the DFT sequence $d_{k'}$, for so long as x is a real function we can exploit the conjugate symmetry of the DFT to construct the second half. The first half is created as follows:

$$d_{k'} = \begin{cases} c_{k'} & k' = 0, 1, 2, \ldots, N^-/2 - 1 \\ 0.5 c_{k'} & k' = N^-/2 \quad (N \text{ even}) \\ c_{k'} & k' = N^-/2 \quad (N \text{ odd}) \\ 0 & k' = N^-/2 + 1, N^-/2 + 2, \ldots, L^-/2, \end{cases} \tag{4}$$

where L^- is defined in the same way as N^- above. The second half of the sequence $d_{k'}$ may then be constructed from the first half by use of

$$d_{L-k'} = d_{k'}^{*} \quad k = 1, 2, \ldots, L^-/2, \tag{5}$$

where * denotes complex conjugation. The factor of 1/2 that appears in Eq. (4) in front of $c_{N/2}$ when N is even and the implicit insertion of this term's complex conjugate in the second half of the sequence are often omitted in hasty implementations of zero-padding interpolation in which sequences of zeroes are simply shoehorned into the middle of the DFT sequence. Fraser discusses the mathematical reason for these nuances, which will prove to be important in establishing the mathematical equivalence of CST and zero-padding for the N even case. Finally then, taking the inverse DFT of the sequence $d_{k'}$,

$$x\left(\frac{2\pi l}{L}\right) = \sum_{k'=0}^{L-1} d_{k'} \exp(j 2\pi l k'/L), \tag{6}$$

yields the more densely sampled sequence $x(2\pi l/L)$, $l = 0, \ldots, L-1$.

2.3. Equivalence of Zero-Padding and CST Interpolation

The most obvious mechanical difference between CST interpolation and zero-padding interpolation is that the former can be used to interpolate a value at a single, arbitrary point θ while the latter necessarily generates simultaneously numerous interpolated points, which lie on a regular grid whose spacing is determined by the zero-padding factor P. However, we will now demonstrate that the values determined by zero-padding on that regular grid match exactly the values that would be obtained from CST interpolation at those points.

We consider the task of using zero-padding to interpolate from N samples, $x(2\pi n/N)$, $n = 0, \ldots, N-1$, to L samples, $x(2\pi l/N)$, $l = 0, \ldots, L-1$, where $L = P \cdot N$ as before. Consider first the case where N is odd. We can divide the sum in Eq. (6) into three segments:

$$x\left(\frac{2\pi l}{L}\right) = \sum_{k'=0}^{(N-1)/2} d_{k'} \exp(j 2\pi l k'/L) + \sum_{k'=(N-1)/2+1}^{L-(N-1)/2-1} d_{k'} \exp(j 2\pi l k'/L) + \sum_{k'=L-(N-1)/2}^{L-1} d_{k'} \exp(j 2\pi l k'/L). \tag{7}$$

Substituting for the $d_{k'}$ in terms of the $c_{k'}$ as specified by Eq. (4), we have

$$x\left(\frac{2\pi l}{L}\right) = \sum_{k'=0}^{(N-1)/2} c_{k'} \exp(j 2\pi l k'/L) + \sum_{k'=L-(N-1)/2}^{L-1} c_{L-k'}^{*} \exp(j 2\pi l k'/L), \tag{8}$$

where the second sum has disappeared because the coefficients $d_{k'}$ are all 0 for k' in this range. Substituting the expression for $c_{k'}$ given by Eq. (3) yields

$$x\left(\frac{2\pi l}{L}\right) = \sum_{k'=0}^{(N-1)/2} \left[\frac{1}{N} \sum_{n=0}^{N-1} x\left(\frac{2\pi n}{N}\right) \exp(-j 2\pi n k'/N)\right] \exp(j 2\pi l k'/L)$$
$$+ \sum_{k'=L-(N-1)/2}^{L-1} \left[\frac{1}{N} \sum_{n=0}^{N-1} x\left(\frac{2\pi n}{N}\right) \exp(j 2\pi n(L-k')/N)\right] \exp(j 2\pi l k'/L). \tag{9}$$

Switching the order of the summations and making the substitution $k'' = k' - L$ in the second term gives

$$x\left(\frac{2\pi l}{L}\right) = \frac{1}{N} \sum_{n=0}^{N-1} x\left(\frac{2\pi n}{N}\right)\left[\sum_{k'=0}^{(N-1)/2} \exp(-j 2\pi n k'/N) \exp(j 2\pi l k'/L) + \sum_{k''=-(N-1)/2}^{-1} \exp(-j 2\pi n k''/N) \exp(j 2\pi l(k'' + L)/L)\right]. \tag{10}$$

The expression $\exp(j2\pi l(k''+L)/L)$ in the second term in brackets is simply equal to $\exp(j2\pi lk''/L)$, so the two terms in brackets may be combined to give

$$x\left(\frac{2\pi l}{L}\right) = \frac{1}{N}\sum_{n=0}^{N-1} x\left(\frac{2\pi n}{N}\right)\left[\sum_{k'=-(N-1)/2}^{(N-1)/2} \exp\left[-j2\pi k'\left(\frac{n}{N}-\frac{l}{L}\right)\right]\right].$$ (11)

At this point we use an identity, which holds for N odd, also invoked by Stark in his derivation of the CST [7]:

$$\frac{1}{N}\sum_{k'=-(N-1)/2}^{(N-1)/2} \exp\left[-j2\pi k'\left(\frac{n}{N}-\frac{l}{L}\right)\right] = \frac{\sin\left[\frac{N}{2}\left(\frac{2\pi l}{L}-\frac{2\pi n}{N}\right)\right]}{N\sin\left[\frac{1}{2}\left(\frac{2\pi l}{L}-\frac{2\pi n}{N}\right)\right]}.$$ (12)

Using this in Eq. (11) yields

$$x\left(\frac{2\pi l}{L}\right) = \sum_{n=0}^{N-1} x\left(\frac{2\pi n}{N}\right)\frac{\sin\left[\frac{N}{2}\left(\frac{2\pi l}{L}-\frac{2\pi n}{N}\right)\right]}{N\sin\left[\frac{1}{2}\left(\frac{2\pi l}{L}-\frac{2\pi n}{N}\right)\right]}.$$ (13)

We see that the values $x(2\pi l/N)$ obtained using zero-padding are exactly those that would be obtained from Eq. (1) for CST interpolation with N odd and $\theta = 2\pi l/L$.

When N is even, the derivation proceeds slightly differently. We now divide the sum in Eq. (6) into five segments:

$$x\left(\frac{2\pi l}{L}\right) = \sum_{k'=0}^{N/2-1} d_{k'} \exp(j2\pi lk'/L) + 0.5d_{N/2} \exp(j2\pi lN/2L) + \sum_{k'=N/2+1}^{L-N/2-1} d_{k'} \exp(j2\pi lk'/L)$$
$$+0.5d_{L-N/2} \exp(j2\pi l(L-N/2)/L) + \sum_{k'=L-N/2+1}^{L-1} d_{k'} \exp(j2\pi lk'/L).$$ (14)

Substituting for the $d_{k'}$ in terms of the $c_{k'}$ as specified by Eq. (4), we have

$$x\left(\frac{2\pi l}{L}\right) = \sum_{k'=0}^{N/2-1} c_{k'} \exp(j2\pi lk'/L) + 0.5c_{N/2} \exp(j2\pi lN/2L)$$
$$+0.5c_{N/2}^* \exp(j2\pi l(L-N/2)/L) + \sum_{k'=L-N/2+1}^{L-1} c_{L-k'}^* \exp(j2\pi lk'/L),$$ (15)

where the third term has disappeared because the coefficients $d_{k'}$ are all 0 for k' in this range. Substituting Eq. (3) for $c_{k'}$ yields

$$x\left(\frac{2\pi l}{L}\right) = \sum_{k'=0}^{N/2-1}\left[\frac{1}{N}\sum_{n=0}^{N-1} x\left(\frac{2\pi n}{N}\right)\exp(-j2\pi nk'/N)\right]\exp(j2\pi lk'/L)$$
$$+0.5\left[\frac{1}{N}\sum_{n=0}^{N-1} x\left(\frac{2\pi n}{N}\right)\exp(-j2\pi n(N/2)/N)\right]\exp(j2\pi lN/2L)$$
$$+0.5\left[\frac{1}{N}\sum_{n=0}^{N-1} x\left(\frac{2\pi n}{N}\right)\exp(j2\pi n(N/2)/N)\right]\exp(j2\pi l(L-N/2)/L)$$
$$+\sum_{k'=L-N/2+1}^{L-1}\left[\frac{1}{N}\sum_{n=0}^{N-1} x\left(\frac{2\pi n}{N}\right)\exp(j2\pi n(L-k')/N)\right]\exp(j2\pi lk'/L).$$ (16)

Now switching the order of the summations, making the substitution $k'' = k' - L$ in the fourth term, and reordering the terms for later convenience allows us to write

$$x\left(\frac{2\pi l}{L}\right) = \frac{1}{N}\sum_{n=0}^{N-1} x\left(\frac{2\pi n}{N}\right)\left[\sum_{k'=0}^{N/2-1}\exp(-j2\pi nk'/N)\exp(j2\pi lk'/L) + \sum_{k''=-(N/2-1)}^{-1}\exp(-j2\pi nk''/N)\exp(j2\pi l(k''+L)/L)\right.$$
$$\left.+0.5\exp(-j\pi n)\exp(j\pi lN/L) + 0.5\exp(j\pi n)\exp(-j\pi lN/L)\right].$$ (17)

The factor $\exp(j2\pi l(k''+L)/L)$ in the second sum is simply equal to $\exp(j2\pi lk''/L)$, so the first two sums in brackets may be combined, as can the third and fourth terms, yielding

$$x\left(\frac{2\pi l}{L}\right) = \frac{1}{N}\sum_{n=0}^{N-1} x\left(\frac{2\pi n}{N}\right)\left[\sum_{k'=-(N/2-1)}^{N/2-1} \exp\left(-j2\pi k'\left(\frac{n}{N}-\frac{l}{L}\right)\right) + \cos\left[\pi N\left(\frac{l}{L}-\frac{n}{N}\right)\right]\right]. \tag{18}$$

We now invoke an identity that holds for N even and whose derivation is similar to that of Eq. (12) [8]:

$$\frac{1}{N}\sum_{k'=-(N/2-1)}^{N/2-1} \exp\left[-j2\pi k'\left(\frac{n}{N}-\frac{l}{L}\right)\right] = \frac{\sin\left[\frac{1}{2}(N-1)\left(\frac{2\pi l}{L}-\frac{2\pi n}{N}\right)\right]}{N\sin\left[\frac{1}{2}\left(\frac{2\pi l}{L}-\frac{2\pi n}{N}\right)\right]}. \tag{19}$$

This allows us to write

$$x\left(\frac{2\pi l}{L}\right) = \sum_{n=0}^{N-1} x\left(\frac{2\pi n}{N}\right)\left[\frac{\sin\left[\frac{1}{2}(N-1)\left(\frac{2\pi l}{L}-\frac{2\pi n}{N}\right)\right]}{N\sin\left[\frac{1}{2}\left(\frac{2\pi l}{L}-\frac{2\pi n}{N}\right)\right]} + \cos\left[\pi N\left(\frac{l}{L}-\frac{n}{N}\right)\right]\right]. \tag{20}$$

Stark and Wengrovitz point out that the second term contributes 0 to $x(2\pi l/N)$, so the final expression is given by

$$x\left(\frac{2\pi l}{L}\right) = \sum_{n=0}^{N-1} x\left(\frac{2\pi n}{N}\right)\left[\frac{\sin\left[\frac{1}{2}(N-1)\left(\frac{2\pi l}{L}-\frac{2\pi n}{N}\right)\right]}{N\sin\left[\frac{1}{2}\left(\frac{2\pi l}{L}-\frac{2\pi n}{N}\right)\right]}\right]. \tag{21}$$

As in the case of N odd, we see that the values $x(2\pi l/N)$ obtained using zero-padding are exactly those that would be obtained from Eq. (2) for CST interpolation with N even and $\theta = 2\pi l/L$.

3. METHODS

3.1. Direct Fourier Image Reconstruction

The goal of two-dimensional tomographic image reconstruction is to estimate a two-dimensional distribution $f(x,y)$ of some object property, such as attenuation coefficient in computed tomography or the concentration of a radionuclide in emission tomography, from a set of one-dimensional projections of the distribution acquired at a number of different projection angles. Formally, the projection at projection angle φ is given by

$$p_\varphi(\xi) = \int_L f(\eta,\xi)d\eta, \tag{22}$$

where $f(\eta,\xi)$ represents the distribution in a coordinate system (η,ξ) rotated from the coordinate system (x,y) by an angle φ, and where L denotes the path of the line of integration. If we denote by $P_\varphi(u)$ the one-dimensional Fourier transform of $p_\varphi(\xi)$ with respect to ξ and by $F(\rho,\varphi)$ the two-dimensional Fourier transform of $f(x,y)$ in polar coordinates, the central slice theorem states that the two are related through

$$P_\varphi(\rho) = F(\rho,\varphi) \qquad \rho \geq 0, \quad 0 < \varphi < 2\pi. \tag{23}$$

Thus the FT of a projection at projection angle φ yields information about the 2D FT of the object along a spoke through the origin of Fourier space oriented at angle φ.

In practice, the projections are measured at a finite number N of angles $\varphi_n = n\pi/N$, $n = 0,...,N-1$, which we assume to be equally spaced over 180°, and each projection, assumed to be of length $2A$, is sampled at a finite number M of projections bins $\xi_m = 2mA/M, m = -(M-1)/2,...,(M-1)/2$. (For notational convenience, we take M to be odd throughout.) When transformed, these projections thus yield a finite set of samples of $F(\rho,\varphi)$, which we denote $F(\rho_m,\varphi_n)$, $\rho_m = m/2A$, $m = 0,...,(M-1)/2$ and $\varphi_n = n\pi/N$, $n = 0,...,2N-1$. Note that the number of angular samples on the polar grid is $2N$ and that the samples span 2π. This reflects the fact that a projection view at angle φ gives two samples of $F(\rho,\varphi)$, one at φ and one at $\varphi + \pi$. Specifically, the polar samples in the range $[0,\pi]$ correspond to the positive frequency

values of the transformed projections, while the polar samples in the range $[\pi, 2\pi]$ can be obtained from the negative frequency values of these transformed projections.

The true challenge in direct Fourier reconstruction is the interpolation from the polar grid above to a Cartesian grid $F(u_i, v_j)$. The size of the interpolated Cartesian grid should be chosen such that none of its points lie beyond the available polar samples; thus it should be the largest square that fits within a circle of radius $(M-1)/4A$. However, in order to ensure that the reconstructed image has the proper scale, this interpolated Cartesian grid should be padded with zeroes to length $(M-1)/A$ prior to performing the inverse 2D FFT. Having established the size of the Cartesian grid in frequency-space units, it must be subdivided into pixels in a way that eliminates or minimizes aliasing in the reconstructed image. To do so, the grid spacing must be no larger than $1/2A$ and preferably an integer divisor smaller.

3.2. Interpolation Between Coordinate Systems

The interpolation of the Cartesian grid values is generally accomplished by scanning the grid point by point, calculating the polar coordinate corresponding to each point and using known polar samples to interpolate a value at that polar point. This last step can be accomplished in many ways. Most crudely, the point could be simply be assigned the value of the nearest polar sample; this is known as nearest-neighbor interpolation. Slightly more accurately, bilinear interpolation could be performed using the four polar samples surrounding the point of interest.

Stark's approach is to perform CST interpolation in the angular direction and truncated Whittaker-Shannon sinc interpolation in the radial direction, obtaining an estimate $\hat{F}(\rho, \varphi)$ of $F(\rho, \varphi)$ at any point (ρ, φ) corresponding to a Cartesian grid point from the measured samples $F(\rho_m, \varphi_n)$ using

$$\hat{F}(\rho, \varphi) = \sum_{m=-(M-1)/2}^{(M-1)/2} \sum_{n=0}^{2N-1} F\left(\frac{m}{2A}, \frac{n\pi}{N}\right) \text{sinc}[2A(\rho - m/2A)] \frac{\sin\left[\frac{1}{2}(N-1)\left(\varphi - \frac{\pi n}{N}\right)\right]}{N \sin\left[\frac{1}{2}\left(\varphi - \frac{\pi n}{N}\right)\right]}. \tag{24}$$

The values of $F(m/2A, n\pi/N)$ for negative m can be determined from the known values for positive m using the relationship $F(-\rho, \varphi) = F^*(\rho, \varphi + \pi)$. This interpolation strategy is valid because the polar samples satisfy two crucial assumptions. First, because the projections are spatially compact, the radial functions on the polar grid should in principle be able to be reconstructed by infinite sinc interpolation from the samples with spacing $\Delta\rho = 1/2A$. The necessary truncation of the series is expected to introduce edge effects leading to high-frequency artifacts in the reconstructed image. Second, the angular functions, while clearly periodic with period 2π, are also expected to be sufficiently bandlimited that they can be sampled in accordance with the Nyquist criterion using a reasonable number of views [10], and can thus be exactly interpolated using the CST.

The reasons that validate Stark's strategy also justify the use of zero-padding interpolation to increase the density of polar samples prior to bilinear interpolation. Because the projections are spatially limited, they can be extended with zeroes to a factor of P times their original lengths prior to Fourier transforming. After transforming, the radial samples of the Fourier transform are P times as dense as previously. At this point, the angular samples at each fixed radial coordinate are subject to a one-dimensional Fourier transform. Because the functions are expected to be bandlimited in the angular direction, each resulting DFT sequence can be zero-padded as discussed in Sect. 2.2 to Q times its original length. Upon performing the inverse DFT, the angular samples of the FT are Q times as dense as previously. At this point interpolation proceeds as in the simple bilinear interpolation scheme discussed above: each Cartesian point is transformed to polar coordinates, and bilinear interpolation is performed among the four surrounding polar points.

3.3. Phantom studies

To compare these three Fourier-domain interpolation approaches, which we will refer to as the bilinear, zero-padding, and CST approaches, we first performed reconstructions of a Shepp-Logan head phantom (see Fig. 1a) in the absence of noise. The projection data consisted of 64 views spanning 180° and each containing 129 projection bins. We reconstructed a 129x129 pixel image (the choice of an odd number of bins and image pixels simplifies the implementation of the direct Fourier algorithms but is by no means necessary). In order to minimize aliasing artifacts, we interpolated onto a Cartesian grid of 257x257 pixels, having the same frequency-space extent as the transforms of the projections yet sampled twice as densely. The central 129x129 pixels of the inverse transform of this array then corresponded to the desired reconstructed image. When zero-padding, we applied zero-padding factors of 2 in both the angular and radial directions, thereby increasing the density of polar samples by a factor of 4.

For each reconstructed image we calculated the normalized root mean square distance of the reconstruction from the true phantom image. This accuracy measure is defined as

$$d = \left(\sum_{i=1}^{R} \sum_{j=1}^{R} (t_{ij} - r_{ij})^2 \middle/ \sum_{i=1}^{R} \sum_{j=1}^{R} \left(t_{ij} - \bar{t} \right)^2 \right)^{1/2} , \tag{25}$$

where t_{ij} and r_{ij} represent the pixel values of the pixel in the ith row and jth column of the true and reconstructed $R \times R$ images, respectively, and $\bar{t}$ denotes the average pixel value in the true image.

To examine the behavior of the algorithms in the face of noise we generated 500 sets of projections of a numerical torso phantom (see Fig. 1b) contaminated with Poisson noise (assuming 100,000 total counts in the sinogram). Each sinogram consisted of 64 projection angles of 65 projection bins spanning 180°. We reconstructed onto a 65x65 pixel array by inverting a Cartesian grid of 129x129 pixels having the same frequency-space extent as the transforms of the projections but sampled twice as densely. For each interpolation technique we calculated the mean of the 500 reconstructions, and the normalized RMS distance of this mean image from the true image. We also computed empirical variance images v_{ij} for each technique using

$$v_{ij} = \frac{1}{N} \left(\sum_{n=1}^{N} (r_{ij}^{(k)})^2 \right) - \left(\frac{1}{N} \sum_{n=1}^{N} r_{ij}^{(k)} \right)^2 , \tag{26}$$

where $r_{ij}^{(k)}$ is the value of the ijth pixel in the kth reconstruction and N is the total number of image realizations.

4. RESULTS AND DISCUSSION

The results of the noise-free reconstructions of the Shepp-Logan head phantom are shown in Fig. 2. Profiles of the row passing through the three small structures near the bottom of the image are also included. The normalized RMS distances and processing times for the three approaches are listed in Table 1. All three algorithms were implemented in Interactive Data Language, though the interpolation step of the circular sampling theorem interpolation was performed by a C subroutine. All computing was performed on an IBM RS6000-based workstation.

Table 1. Normalized RMS distances and processing times for reconstruction of the noise-free Shepp-Logan phantom.

Interpolation Method	Normalized RMS Distance	Processing Time
Bilinear	0.268	7.28 s
Zero-Padding	0.126	8.53 s
CST	0.148	2511.41 s

It is apparent, both qualitatively and quantitatively, that zero-padding interpolation produces images of comparable or superior quality to CST at a fraction of the computational cost. To be fair, CST can be made to run more rapidly by constraining the convolutions to smaller neighborhoods around the points of interest, but this comes at some cost in accuracy, and computing time still would barely approach that of the zero-padding approach, which benefits from the famed speed of the FFT. The excellent performance of the zero-padding approach relative to CST is explained in large part by their equivalent exactness in performing the angular interpolation and their equivalent approximations in performing the radial interpolation. So long as the polar samples satisfy periodicity and sampling criteria in the angular direction, the values interpolated in that direction by both methods are exact. In the radial direction, where the samples are not of a periodic function, the approximation entailed in using zero-padding interpolation is comparable to that involved in truncating the Whittaker-Shannon sinc interpolation series in Stark's CST approach. The error introduced in using bilinear interpolation in the final stage of the zero-padding approach can of course be made negligibly small by choosing sufficiently large zero-padding factors, though in practice, factors as small as 2 in each direction produce excellent results.

The accuracy of bilinear interpolation alone is, not surprisingly, worse than either of the other two methods. The reconstruction suffers from a depression in values toward the edges of the image, as can clearly be seen in the line profile. This rolloff phenomenon arises because of the deviation of the effective interpolation kernel from an ideal 2D-sinc interpolator—which the CST and zero-padding approaches approximate much more closely—and is discussed in greater detail by Jackson *et al.* [3].

The mean images of the noise-contaminated torso phantoms are illustrated in Fig. 3, along with line profiles passing through the bright annulus representing the heart near the top of the image. The normalized RMS distances listed in Table 2 demonstrate the same trends as in the noise-free reconstruction just discussed. Fig. 4 depicts the variance maps for the three approaches, along with profiles along the same line as illustrated in Fig. 3. The total variance in each map is listed in Table 2, where it can be seen that simple bilinear interpolation leads to a lower total empirical variance than the other two, whose

results are comparable. This is not surprising as the relatively crude bilinear interpolation smooths over noisy variations in Fourier space more than the CST and zero-padding approaches.

Table 2. Normalized RMS distances of mean images and total empirical variances for 500 reconstructions of a torso phantom containing Poisson noise (100,000 counts).

Interpolation Method	Normalized RMS Distance	Total Variance
Bilinear	0.252	2.78×10^6
Zero-Padding	0.122	4.16×10^6
CST	0.127	4.79×10^6

5. CONCLUSIONS

We have examined the use of three different methods for performing the interpolation from a polar to a Cartesian grid that is encountered in direct Fourier image reconstruction. Simple bilinear interpolation is, not surprisingly, found to yield poor-quality images, while a relatively efficient hybrid method, in which zero-padding interpolation is used to increase the density of polar samples prior to bilinear interpolation onto the Cartesian grid is found to yield comparable results to the much more computationally intense circular sampling theorem approach. We have attempted to account for the success of this hybrid approach in three ways. First, we have proved that zero-padding interpolation is exact and equivalent to circular sampling theorem interpolation when used to interpolate periodic functions sampled in accordance with the Nyquist criterion, which are precisely the sort that arise in the angular dimension of the polar grid. Second, we have pointed out that both methods make equivalent approximations in interpolating in the radial direction. Finally, we have indicated that the error arising in the subsequent use of bilinear interpolation in the hybrid approach can be minimized by increasing the density of polar samples sufficiently through zero-padding.

ACKNOWLEDGMENTS

REFERENCES

1. G. T. Herman, *Image Reconstruction from Projections,* Academic Press, Orlando, 1980.
2. J. D. O'Sullivan, "A fast sinc function gridding algorithm for Fourier inversion in computed tomography," *IEEE Trans. Med. Imaging* **4,** pp. 200-207, 1985.
3. J. I. Jackson, C. H. Meyer, D. G. Nishimura, and A. Macovski, "Selection of a convolution function for Fourier inversion using gridding," *IEEE Trans. Med. Imaging* **10,** pp. 473-478, 1991.
4. H. Stark, J. W. Woods, I. Paul, and R. Hingorani, "Direct Fourier reconstruction in computer tomography," *IEEE Trans. Acoust., Speech, Signal Processing* **29,** pp. 237-245, 1981.
5. H. Stark, J. W. Woods, I. Paul, and R. Hingorani, "An investigation of computerized tomography by direct Fourier inversion and optimum interpolation," *IEEE Trans. Biomed. Eng.* **28,** pp. 496-505, 1981.
6. S. X. Pan and A. C. Kak, "A computational study of reconstruction algorithms for diffraction tomography: interpolation versus filtered backpropagation," *IEEE Trans. Acoust., Speech, Signal Processing* **31,** pp. 1262-1275, 1983.
7. H. Stark, "Sampling theorems in polar coordinates," *J. Opt. Soc. Am.* **69,** pp. 1519-1525, 1979.
8. H. Stark and M. Wengrovitz, "Comments and corrections on the use of polar sampling theorems in CT," *IEEE Trans. Acoust., Speech, Signal Processing* **31,** pp. 1329-1331, 1983.
9. D. Fraser, "Interpolation by the FFT revisited—an experimental investigation," *IEEE Trans. Acoust., Speech, Signal Processing* **37,** pp. 665-675, 1989.
10. X. Pan, "Quasi band-limited properties of Radon transforms and their implications for increasing angular sampling densities," submitted to *IEEE Trans. Med. Imaging.*
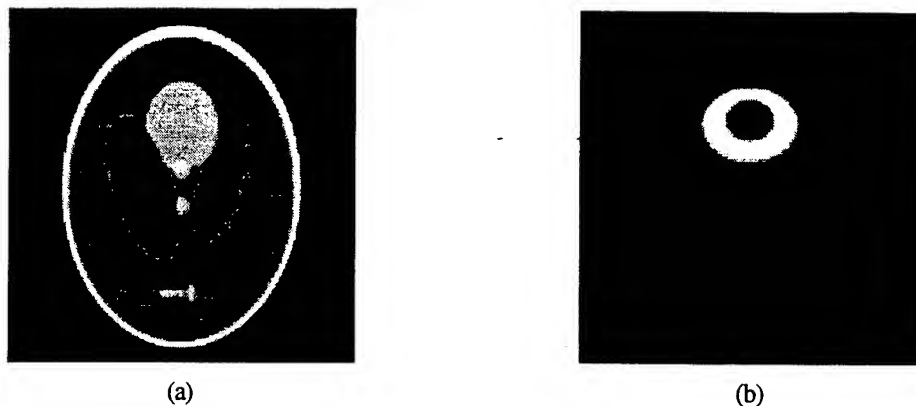
Figure 1. Original numerical phantoms used in noise-free (a) and noise-contaminated (b) studies.
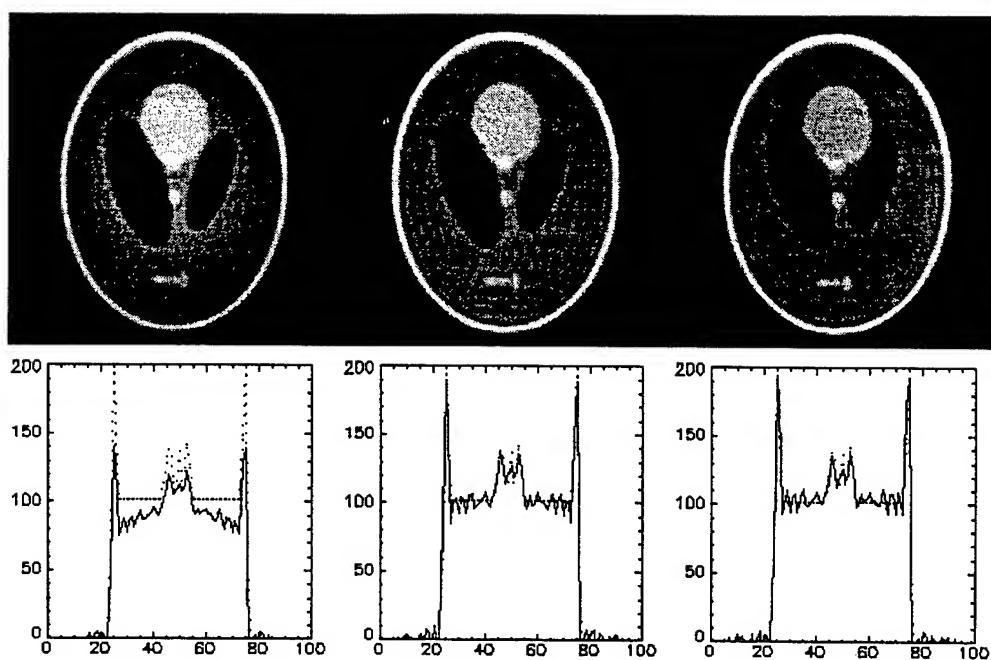

Figure 2. Direct Fourier reconstructions of the noise-free numerical Shepp-Logan phantom using simple bilinear interpolation (left), zero-padding interpolation followed by bilinear interpolation (center), with zero-padding factor 2 in both the angular and radial directions, and circular sampling theorem interpolation (right). The line profiles pass through the three small structures near the bottom of the images, with the dotted line representing the true values.
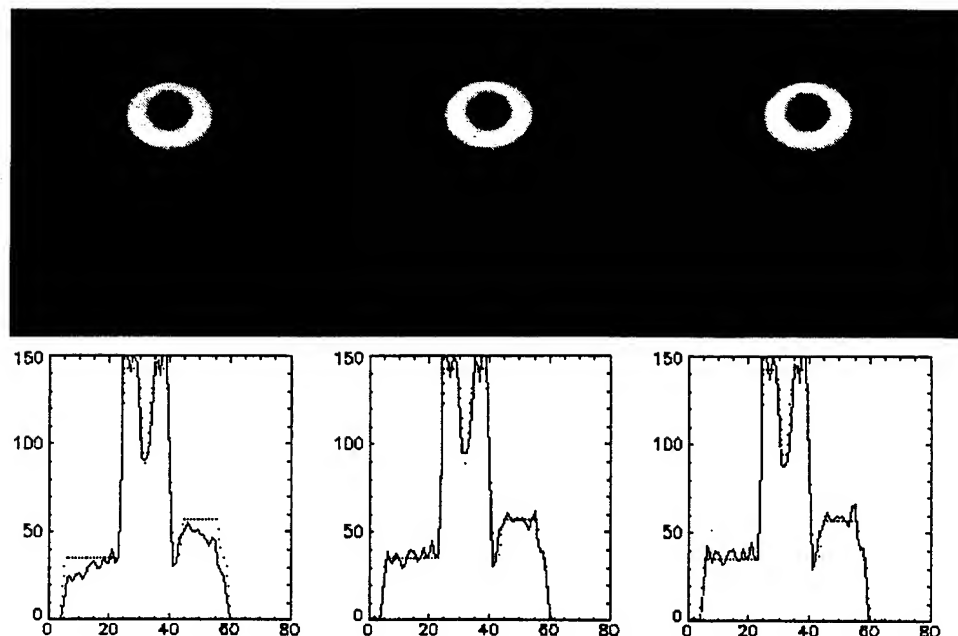
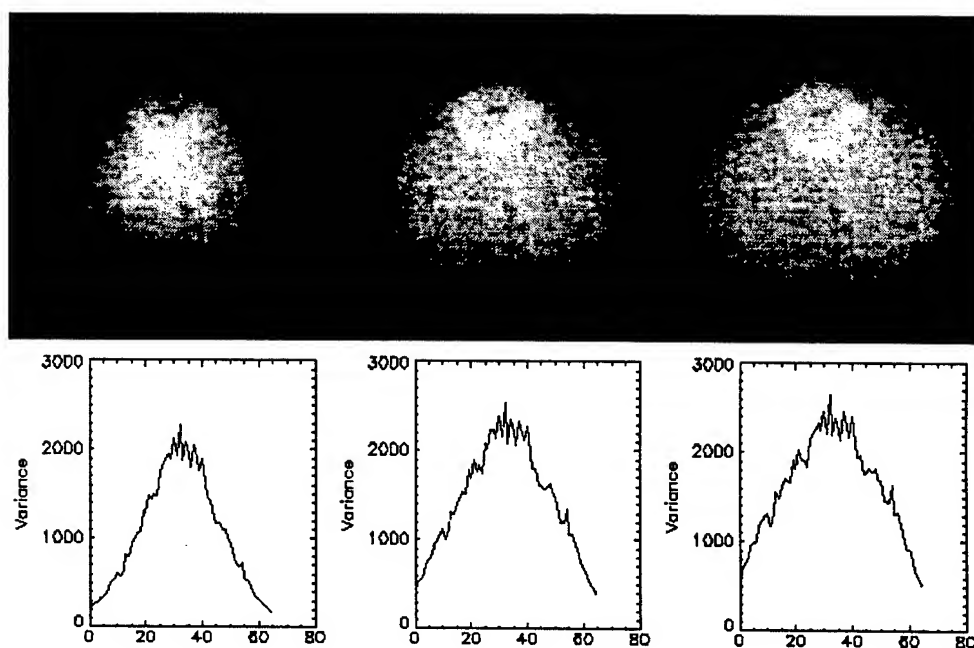**Figure 3.** Means of 500 direct Fourier reconstructions of a numerical torso phantom contaminated with Poisson noise (100,000 counts) using bilinear interpolation (left), zero-padding interpolation followed by bilinear interpolation (center), with zero-padding factor 2 in both the angular and radial directions, and circular sampling theorem interpolation (right). The line profiles pass through the annulus near the top of the images, with the dotted line representing the true values.

Figure 4. Variance images of 500 direct Fourier reconstructions of a numerical torso phantom contaminated with Poisson noise (100,000 counts) using bilinear interpolation (left), zero-padding interpolation followed by bilinear interpolation (center), with zero-padding factor =2 in both the angular and radial directions, and circular sampling theorem interpolation (right). The line profiles pass through the annulus near the top of the images.

Efficient sinogram smoothing for dynamic neuroreceptor PET imaging

X. Pan, P.J. La Rivière, J. Ye*, J. Mukherjee, and C.-T. Chen

Department of Radiology
*Graduate School of Business
The University of Chicago, Chicago, IL 60637

ABSTRACT

We have developed image-restoration techniques applicable to dynamic positron emission tomography (PET) that improve the visual quality and quantitative accuracy of neuroreceptor images. Starting with data from a study of dopamine D-2 receptors in rhesus monkey striata using selective radioligands such as [F-18]fallypride, we performed a novel effective three-dimensional smoothing of the dynamic sinogram at a much lower computational cost than a truly three-dimensional, adaptive smoothing. The processed sinogram was then input to a standard filtered back-projection algorithm and the resulting images were sharper and less noisy than images reconstructed from the unprocessed sinogram. Simulations were performed and the radioligand binding curves extracted from the restored images were found to be smoother and more accurate than those extracted from the unprocessed reconstructions. Comparison was also made to reconstructions from sinograms processed by the principal component analysis/projection onto convex sets (PCA-POCS) algorithm.

Keywords: PET, image restoration, neuroreceptors, dopamine, tomographic reconstruction.

1. INTRODUCTION

Due to their involvement in neurochemical functions as well as various pathophysiologies, neuroreceptors have become important candidates for study by non-invasive imaging techniques such as positron emission tomography (PET). The dopaminergic neurotransmitter receptor system in particular has been related to human brain disorders on the grounds that dopamine-replacement therapy can alleviate Parkinson's disease, and that many antipsychotic drugs are dopamine receptor antagonists. The dopamine neurotransmitter receptor hypothesis has received a great deal of attention as a biochemical model for schizophrenia, tardive dyskinesia, dystonia, Parkinson's disease, Huntington's disease, and alcoholism. We have developed methods to study dopamine D-2 receptors in the monkey striata using selective radioligands such as [F-18]fallypride and [F-18]desmethoxyfallypride.[1,2] However, due to the low spatial resolution and other shortcomings of PET imaging, identification and quantitation of these receptor sites within small regions of the brain is a significant challenge. Image-restoration techniques that correct or compensate for some of these shortcomings would improve the accuracy of the measurement of the concentration of radioligand in various brain regions.

The shortcomings of PET imaging are numerous and well-known. The coincidence line data that emerge from PET studies are invariably modified by photon attenuation, contaminated by scatter and accidentals, blurred due to finite detector size, and degraded by noise. It is possible to incorporate corrections for such flaws directly into the reconstruction algorithm, but an alternative approach calls for processing the sinogram itself to compensate for such effects prior to reconstruction by a standard method such as filtered back-projection (FBP). For instance, a technique known as projection onto convex sets (POCS) has been developed to compensate for the blur caused by finite detector size, while a technique known as principal component analysis (PCA) can provide noise suppression in the temporal direction.[3-6]

In principle, a fully three-dimensional, adaptive smoothing of the sinogram would make maximal use of the correlations present in the data and thus provide optimal noise suppression. By adaptive smoothing, we mean that the degree of smoothing is determined adaptively from the statistics inherent in subsets of the data and generally varies from subset to subset. This is in contrast to a simple smoothing kernel, for example, which is applied globally to the whole data set. Unfortunately, fully three-dimensional, adaptive smoothing algorithms are not readily available, and would in any case be very expensive computationally. By exploiting the convolution-multiplication theorem of Fourier analysis, however, we have devised a method that provides *effective* three-dimensional, adaptive smoothing at the expense of a two-dimensional Fourier transform and its inverse and a one-dimensional, adaptive smoothing operation.

2. METHODOLOGY

We performed a dopamine D-2 neuroreceptor-imaging study of the brain of a rhesus monkey on the PETT VI system at the University of Chicago. The radioligand used was F-18-labeled fallypride, developed for specific dopamine D-2 receptor binding by Mukherjee.[1,2] Five slices were obtained; in this paper, we will be concerned with the slice passing through the striata, shown in Figures 1-6. After the injection of 2.45 mCi of ^{18}F-fallypride, we performed two consecutive 22-frame scans, with each frame lasting 2.62 seconds. The frames contained approximately 360,000 counts at the beginning of the scan and about 60,000 counts at the end. To evaluate fallypride binding specificity to 5HT-2 and dopamine D-2 receptors, competition studies were performed by injecting ketanserin (which competes for 5HT-2) and raclopride (which competes for D-2) at 70 and 100 minutes after the injection of fallypride. Accidentals and scatters were not corrected for; photon attenuation was compensated by use of a transmission scan (containing 6.4 million counts) acquired with a Ge-68/Ga-68 ring-source prior to the administration of the radiotracer. The emission scan was acquired using the PETT VI high-resolution mode. After acquisition, the raw data were rebinned into a 72 angle x 108 bin x 44 time slice sinogram.

The extracted sinogram is blurred and generally quite noisy, and direct FBP reconstruction leads to correspondingly blurred and noisy images (see Figures 1 and 2). The sinograms that we use as input to our method are first processed by using a projection onto convex sets (POCS) signal-recovery technique that corrects for the effect of blur due to finite detector size. Our technique then suppresses noise by means of an effective three-dimensional, adaptive smoothing achieved at the cost of a two-dimensional Fourier transform, its inverse, and a one-dimensional, adaptive smoothing. Starting with an n-dimensional data set, it can be shown that the following sequence of operations leads to an effective n-dimensional smoothing of the data: (1) an (n-1)-dimensional Fourier transform; (2) a one-dimensional, adaptive smoothing over the untransformed parameter; and (3) an inverse (n-1)-dimensional Fourier transform.[7] In the study at hand, the dynamic sinogram is a function of three variables: time, angle, and bin number. We first perform a two-dimensional Fourier transform of this sinogram over the time and angle variables. We can regard the outcome of this transform as a set of one-dimensional functions of the remaining variable—bin number—each labeled by a pair of temporal and angular frequency values. It is these one-dimensional functions of bin number that we smooth in a manner to be described below. We then take an inverse two-dimensional Fourier transform to recover the sinogram, which is effectively smoothed over all three variables.

The smoothing technique we apply to each of these one-dimensional functions of bin number is known as penalized least-squares smoothing, in which the smoothed curve $\hat{g}(\xi)$ is the minimizer of the functional:

$$S(g) = \sum_{i=1}^{N} \left\{ Y_i - g(\xi_i) \right\}^2 + \alpha \int_a^b \left\{ g''(\xi) \right\}^2 d\xi \ , \tag{1}$$

where N is the number of data points, ξ_i is the abscissa value of the ith data point, Y_i is the corresponding ordinate value, a and b are the limits of the interval on which the data are collected, and $g(\xi)$ is a twice-differentiable function representing the estimated curve. The two terms in the functional represent the competing goals of achieving a good fit to the data while maintaining a smooth curve, with the parameter α mediating the tradeoff. For instance, if α is zero, the smoothness constraint disappears and the minimizing curve will be a piecewise linear interpolant to the data; if α grows large, the smoothness constraint dominates and the curve approaches a simple linear fit to the data. For intermediate values of α, the minimizing curve balances the goodness-of-fit and smoothness constraints. It can in fact be shown that the minimizers of this functional $S(g)$ will always be members of a class of functions known as natural cubic splines.[8] These are piecewise polynomial curves that join at the abscissa values ξ_i, where they are continuous up to and including the second derivative.

Clearly the choice of α will greatly influence the shape of the spline fit to the data. One could, in principle, choose a single value of α to be applied to all of the one-dimensional functions of ξ, but it is in fact preferable to allow the statistics of the data itself to dictate the choice of α separately for each such function. This adaptiveness is achieved through an algorithm known as cross-validation, which searches for the α which generates the curve that best predicts the results of new observations. Given that one does not necessarily have ready access to new observations, such an observation is actually generated by omitting one data point from the data set. One then fits a curve to the remaining points and compares the fit curve's prediction for the omitted data point to the actual value of the omitted data point. In practice, a mean measure of predictive accuracy for a given value of α is determined by leaving out each data point in turn and summing the mean squared errors of the ensuing predictions. The α that yields the lowest such mean-squared error is selected. The procedure as described is computationally intensive, but in practice the same end can be achieved approximately with an analytic and efficient solution known as generalized cross-validation.[7]

This smoothing algorithm was applied to the monkey data described above and the quality of the resulting

reconstructions can be assessed visually as we discuss in the next section. However, because the true radioligand concentrations in various areas of the brain for various time slices are not known, it was necessary to generate a simulated data set in order to evaluate the quantitative accuracy of the reconstruction algorithms. The noise-free simulation of dopamine D-2 receptor sites in the rhesus monkey consisted in two small ellipses (simulating the striatal structures) located within a large ellipse (representing the monkey brain). The major and minor axes of the small ellipses were 20 mm and 8 mm, respectively. Dopamine D-2 receptor sites were assumed to exist exclusively in the striatal regions. A 48-frame dynamic sequence of fallypride-binding images was generated by the use of a nonspecific-binding time sequence for the brain plus a specific-binding sequence for the striata. Noisy PET data were then simulated assuming the PETT VI high-resolution mode geometry. The simulated noisy data contained roughly 100,000 counts in the first frame and 16,000 counts in the last frame.

3. RESULTS

Six sets of reconstructed images are shown in Figures 1-6. Figures 1 and 2 depict, respectively, the results of reconstructing real and simulated monkey striata PET data which has simply been rebinned into a sinogram with no further processing. The reconstruction algorithm is simple FBP with a ramp filter. The images are both noisy and blurry; the striata are simply not visible in the early time slices and they are not sharply resolved in the later time slices.

Figures 3 and 4 depict images reconstructed from the same real and simulated data which has now been processed by the aforementioned PCA-POCS algorithm prior to reconstruction by FBP with a ramp filter. The PCA stage of the algorithm effects a strong smoothing in the temporal direction by performing a mathematical transform of the data that allows extraction of the principal signal-bearing components of the data while suppressing the noise-bearing components. The POCS stage of the algorithm compensates for various degradations of the data, such as blurring due to finite detector size, while allowing the imposition of additional constraints such as non-negativity of the data. As expected, we observe a marked improvement in resolution compared to Figures 1 and 2 as well as a noticeable change in the noise structures present.

Finally, Figures 5 and 6 illustrate the results of reconstructing from real and simulated data that have been processed by the POCS algorithm followed by our effective three-dimensional smoothing procedure. The reconstruction algorithm is again FBP with a ramp filter. The resulting images are comparable in sharpness to those reconstructed with PCA-POCS processing but with noticeably improved contrast in the earliest time slices, where the striata can be clearly identified.

After reconstruction, the simulated images were also subject to ROI quantitation analysis to evaluate the quantitative accuracy of the various reconstruction schemes. The calculated binding curves as well as the true binding curves are shown in Figure 7. The reconstruction from the unprocessed sinogram results in a severe underestimate of the radioligand concentration. Reconstruction from the POCS processed sinogram provides a more accurate curve, though with a large variance. Applying the effective three-dimensional smoothing to the POCS-processed sinogram reduces the variance of the extracted binding curve. The most accurate curve is the one extracted from the PCA-POCS-processed images, which is both relatively smooth and in good agreement with the true curve for all time points.

4. CONCLUSIONS

The results from the combination of POCS with our three-dimensional smoothing algorithm are encouraging. While the specific binding curve computed from these simulated images appears to underestimate the true value, the curve extracted from images processed by POCS alone similarly underestimates the specific binding, which suggests that the POCS step of the algorithm may be the source of that discrepancy. The fact that this underestimate does not arise in the PCA-POCS case may stem from the fact that PCA is applied prior to POCS processing in this case. This sequence, in which the resolution recovery step is applied after the smoothing step, may diminish the influence of the partial volume effect. This suggests that application of the three-dimensional smoothing prior to POCS processing may bear examination in future research.

These successes of the effective three-dimensional adaptive smoothing algorithm are achieved at a reasonable computational cost compared to truly three-dimensional adaptive smoothing. The computational burden of the effective smoothing operation is on the order of 100 seconds for a 108 bin x 72 angle x 44 time slice data set processed on a Silicon Graphics Indy 2 workstation, while a fully three-dimensional approach could take an hour or more on a comparable system.

Finally, it should be noted that the effective smoothing algorithm is by no means limited to application in three dimensions. For static PET or SPECT imaging a similar algorithm can achieve an effective two dimensional smoothing of the sinogram at the cost of a one-dimensional Fourier transform and its inverse in addition to a one-dimensional smoothing operation.[10] Likewise, for dynamic imaging comprising a fully three-dimensional data set, an effective four-dimensional smoothing can be achieved in an analogous manner.

ACKNOWLEDGMENTS

REFERENCES

1. J. Mukherjee, Z.Y. Yang, M.K. Das, and T. Brown, "Fluorinated benzamide neuroleptics 3. Development of (S)-N-[(1-allyl-2-pyrrolidinyl)-methyl]-5-(3[F-18]-fluoropropyl)-2,3-dimethoxybenzamide as an improved dopamine D-2 receptor tracer", *Nucl. Med. Biol.* **22**, pp. 283-296, 1995.

2. J. Mukherjee, Z.Y. Yang, T. Brown, J. Roemer, and M. Cooper, "18F-Desmethoxyfallypride: A fluorine-18 radiotracer with properties similar to carbon-11 raclopride for PET studies of dopamine D-2 receptors", *Life Sciences* **59**, pp. 669-678, 1996.

3. H. Stark and P. Oskoui, "High-resolution image recovery from image-plane arrays, using convex projections", *J. Opt. Soc. Am. A* **6**, pp. 1715-1726, 1989.

4. C.-M. Kao, J.T. Yap, and M.N. Wernick, "Improving region-of-interest quantitation in dynamic PET studies by principal component analysis and sinogram restoration", *JNM* **36**, p. 185p, 1995.

5. C.-M. Kao, J.T. Yap, J. Mukherjee, and M.N. Wernick, "Improving region-of-interest quantitation in dynamic PET studies by low-order approximation and pre-reconstruction sinogram restoration", submitted to *IEEE Trans. MI.*

6. C.-M. Kao, "Pre-reconstruction sinogram restoration for positron emission tomography", Ph.D. Dissertation, Univ. of Chicago, 1997.

7. X. Pan, "A novel approach for n-dimensional smoothing", FMI / Department of Radiology Technical Report, 1996.

8. P.J. Green and B.W. Silverman, *Nonparametric Regression and Generalized Linear Models*, pp. 11-27, Chapman & Hall, London, 1994.

9. G. Wahba, *Spline Models for Observational Data.* SIAM CBMS-NSF, Philadelphia, 1990.

10. P.J. La Rivière, X. Pan, B.C. Penney, C.-T. Chen, and R. Reba, "Few-angle cardiac SPECT with spline processing: initial results and radial-extent limitations", submitted to SNM Annual Meeting, 1997.

Further author information —

X.P.(correspondence): Email: x-pan@uchicago.edu; Telephone: (773) 702-1293
P.J.L.: Email: pjlarivi@midway.uchicago.edu; Telephone: (773) 702-0696
J.M.: Email: jogm@fciad3.bsd.uchicago.edu; Telephone: (773) 702-3226
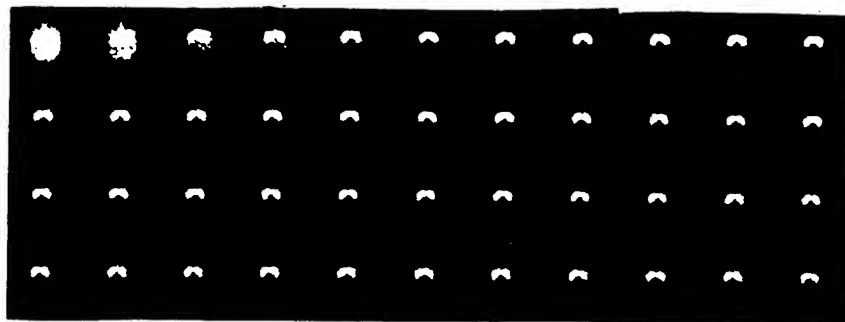C.C.: Email: c-chen@uchicago.edu; Telephone: (773) 702-6269

Figure 1: FBP reconstruction of unprocessed sinogram data from actual D-2 dopamine receptor PET study. The striata are not clearly visible in the first few time slices and are poorly resolved throughout the sequence.
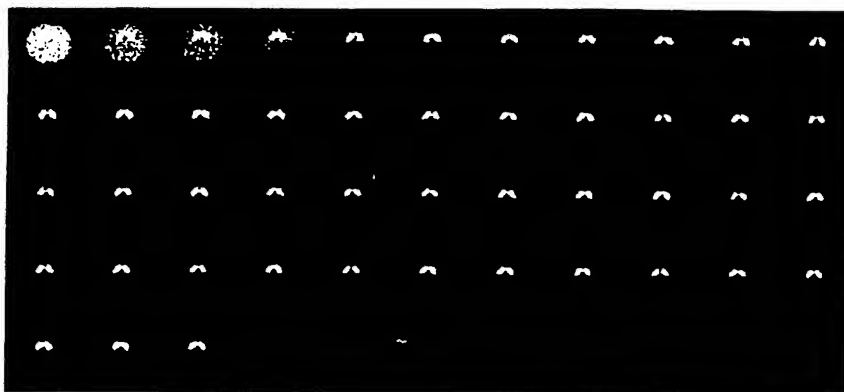

Figure 2: FBP reconstruction of unprocessed sinogram data from simulated D-2 dopamine receptor PET study. The striata are not clearly visible in the first few time slices and are poorly resolved throughout. Note also the severe noise structures in many of the time slices.
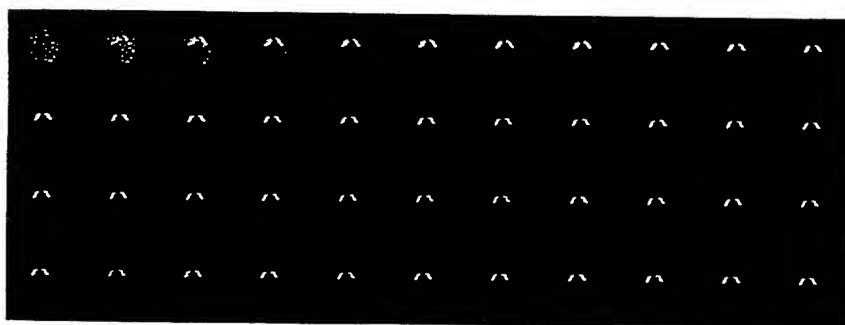

Figure 3: FBP reconstruction of PCA-POCS processed sinogram data from actual D-2 dopamine receptor PET study. The resolution of the striata is improved throughout relative to the unprocessed reconstructions in Figures 1 and 2, but it is still difficult to discern the striata in the early time slices.
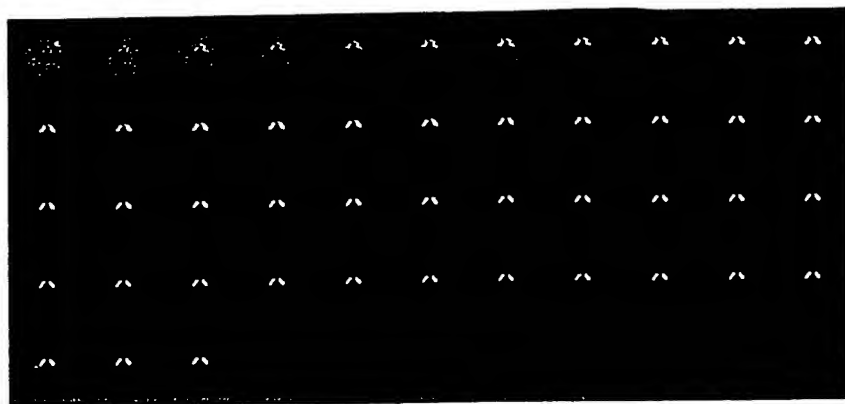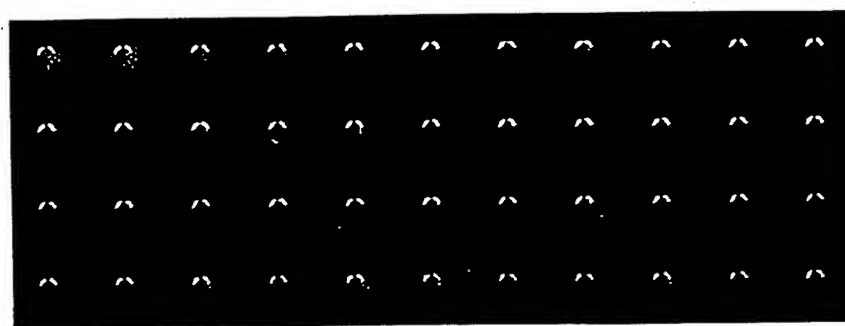
Figure 4: FBP reconstruction of PCA-POCS processed sinogram data from simulated D-2 dopamine receptor PET study. The resolution of the striata is improved throughout relative to the unprocessed reconstructions in Figures 1 and 2, but it is still difficult to discern the striata in the early time slices.

Figure 5: FBP reconstruction of sinogram data from actual D-2 dopamine receptor PET study that have been processed by POCS and the three-dimensional smoothing algorithm described in the text. The resolution of the striata throughout is comparable to that of the PCA-POCS images, while the contrast in the early slices is clearly superior, for the striata are now readily visible.
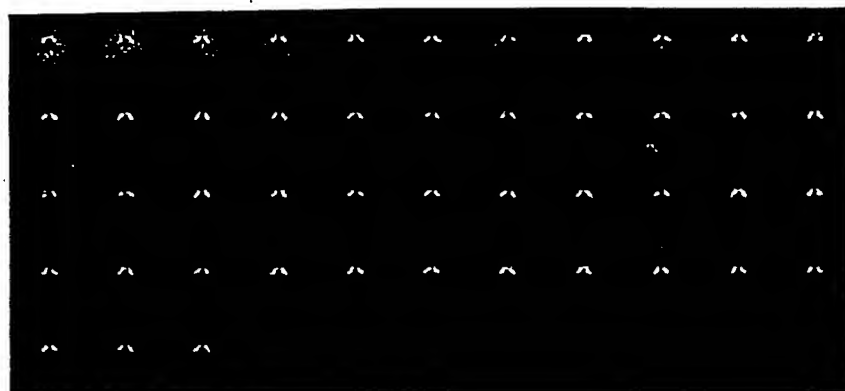

Figure 6: FBP reconstruction of sinogram data from simulated D-2 dopamine receptor PET study that have been processed by POCS and the three-dimensional smoothing algorithm described in the text. The resolution of the striata throughout is comparable to that of the PCA-POCS images, while the contrast in the early slices is clearly superior, for the striata are now readily visible.
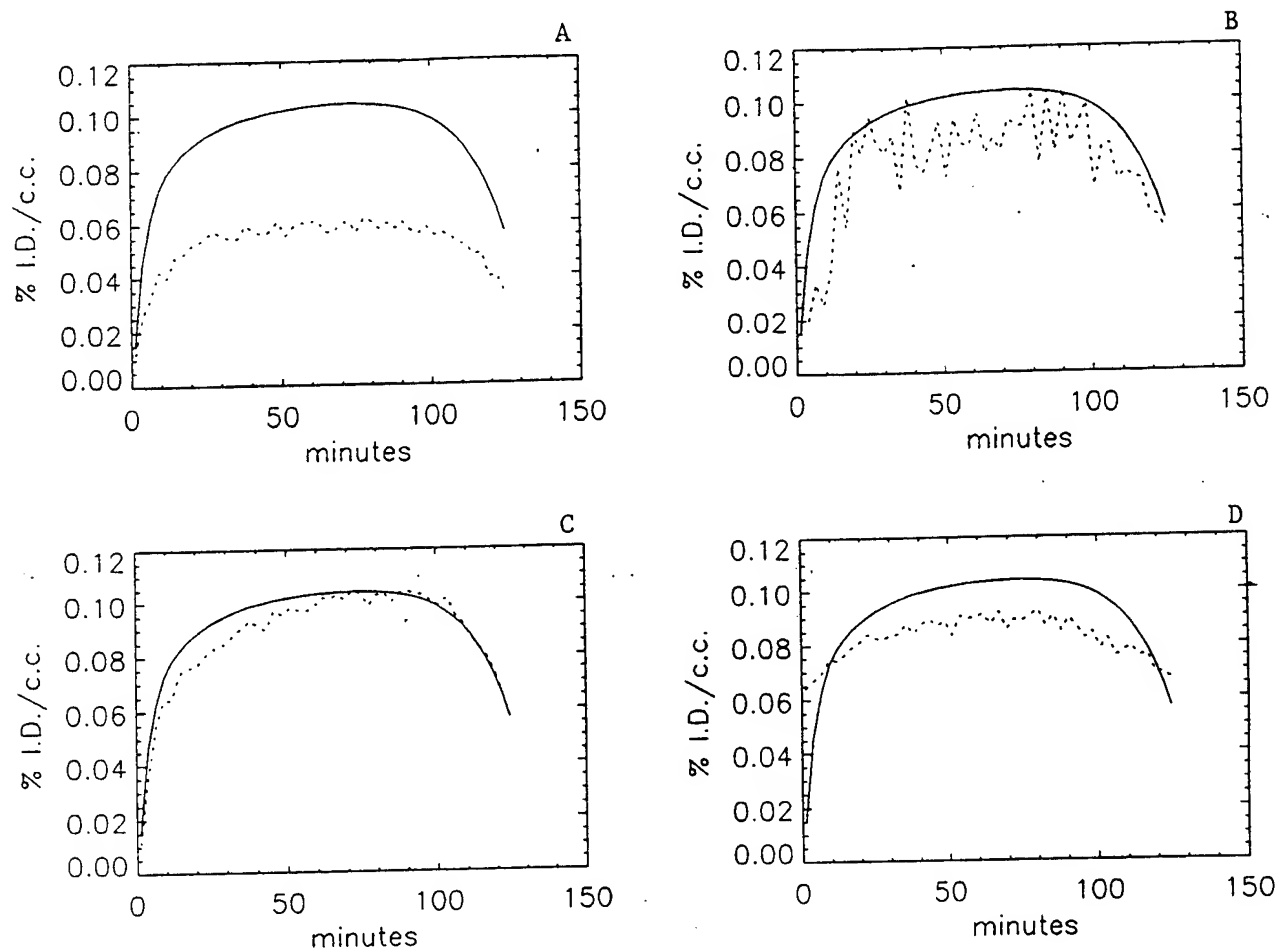
Figure 7: Fallypride binding curves obtained from simulated PET image sequences using four different sinogram processing techniques prior to reconstruction: (a) no processing; (b) POCS alone; (c) PCA-POCS; and (d) POCS with three-dimensional smoothing. The solid curve in each figure is the true specific binding curve. The unprocessed reconstruction results in a severe underestimation of specific binding. The POCS result is closer to the true result, but the curve is very noisy. The PCA-POCS agrees very well with the true result and is much smoother than the POCS curve. Finally, the POCS with three-dimensional smoothing curve is smooth and generally in good agreement with the true curve, though slightly underestimated.